

Materials for sustainability

Whether in energy generation or environmental protection, materials research has already made many contributions. But the community has further to go to reduce the impacts of entire cycles of materials use.

Researchers working on materials have played an active role in cleaning up the environment over the past 30 years. Materials scientists have improved the quality of released effluent and exhaust plumes, the use of catalysts to avoid unwanted by-products in chemical processes, and the treatment of waste. They have moved from solvent-based coatings to powder coatings and biodegradable materials, and are helping to generate cleaner energy. For example, they are developing photovoltaic materials with improved conversion efficiencies, and materials that form components of the various types of fuel cell under development.

But those designing materials for such purposes should weigh up environmental impact as well as material functionality, and both the costs and benefits of synthesizing and processing materials. There also needs to be a shift in the materials community at large towards technology sustainability, specifically in terms of energy and environmental impacts. For example, it has been estimated that solar cells take between three and eight years to pay back their energy costs. Significant energy input stems from the aluminium or steel frames in which the cells are placed. Additional costs come from the manufacture of solar cells, the disassembly and recycling of components, and finally from chemicals that might pose occupational health risks to workers and are difficult to dispose of. It could be argued that the energy delivered and the lack of carbon dioxide produced eventually outweighs these costs. But careful analysis of the life-cycle assessment of solar cells will pinpoint areas of improvement in material and device design to maximize the environmental benefit.

Metal recycling is an important aspect of sustainability. The mining and processing of aluminium ore yields large quantities of carbon dioxide, oxides of sulphur and nitrogen, and volatile organic compounds, not to mention waste for disposal in landfill sites and the large-scale consumption of water. Research is required to ensure that, by 2008, there is a significant amount of recycled aluminium in use. The mining and production of virgin aluminium should be reduced considerably, although elimination is sure to be a distant goal.

Meeting of minds

Industrial companies must address the sustainability of materials and technologies to help ensure their long-term survival. At the First Materials Science Forum for Sustainable Technologies (MATFORUM) last month at the University of Augsburg in Germany, some of the world's largest manufacturing companies — including Ford, Schott Glass, Pfizer, Motorola and TotalFinaElf — came head to head with members of the World Wide Fund for Nature, the United Nations, the US Environmental Protection Agency and others. The meeting presented an opportunity to discuss complex and interesting sustainability problems in technology, and to find solutions.

Too many presentations were simply company advertisements or a defence of the industry under discussion. But there were also significant moments in the meeting when the presenters were honest and open, and these led to some productive discussions.

Delegates were invited to embrace strategies directed towards 'de-materialization'. Here, production and consumption cycles of materials and devices are designed to be closed loops, where the goals

are the extraction of the energy content of waste, the recycling of materials, and the re-use and re-integration of components. Other discussions focused on misuse or misunderstanding of terminology, such as 'recycling'. Formally, this is a process in which material is returned to a previous stage of a cyclic process, so it must be re-usable in the same function that it had before being re-processed. This is not to be confused with downcycling, in which a used material is processed for a lower-technology application, shifting the disposal problem further downstream but not removing it.

There were many university researchers at MATFORUM from the traditional fields of environmental materials science (such as recyclability and clean energy), but the broader academic materials research community missed a significant opportunity to apply their knowledge and experience to solving some of these sustainability problems.

New technologies

Researchers must also help to ensure that new problems are not created. After all, the likely long-term commercial success of technologies being conceived in the lab now could depend on their sustainability properties, as well as their functional properties. This may sound obvious, but how many researchers in the materials community, other than those involved directly in sustainable technologies, really pay due attention to such notions?

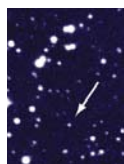
It is time for the broader materials community to think through the environmental implications of its work, and even to consider taking on the demanding challenges placed on industry. These are exciting projects, presenting new problems to be solved in many well-established materials disciplines, such as steel and cement technologies, as well as high-technology and novel materials.

Some universities and government initiatives are already fostering new thinking. Undergraduate materials courses are including consideration of the environmental implications of materials manufacture. The European Commission places sustainability as one of the main technological objectives in its "New Materials and Production Technologies" themes. The US National Science Foundation and Environmental Protection Agency have formed a partnership, providing funds for fundamental research in the physical sciences, which, it is hoped, will lead to the discovery, development and evaluation of new industrial methods that are environmentally benign.

In a very different context, materials researchers in Africa are slowly becoming more active globally, with the African Materials Research Society due for launch in December. Perhaps these researchers will be instrumental in the pursuit of sustainable technologies.

But convincing everyone in the materials community of the value of this shift in perspective is likely to be a slow process. One tangible outcome of MATFORUM is the preparation of an Augsburg declaration (under revision at the time of going to press; see http://www.amu-augsburg.de/matforum/augsburg_materials_declaration.pdf), which asks scientists to consider all of the sustainability issues arising from the large-scale manufacture of their materials. As the old English proverb has it, fine words butter no parsnips. It is to be hoped that more materials researchers will rise to the challenges embodied in the declaration. ■

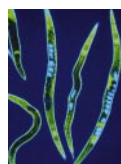




Rock star
Astronomers net
largest Kuiper-belt
object
p546



Trial run
Switzerland lays
down the law on
transgenic crops
p547



Nematode Nobel
Worm turns into
physiology prize
for geneticists
p548



Ritual murder
Archaeologists
unearth victims of
ancient slaughter
p551

Regulators split on gene therapy as patient shows signs of cancer

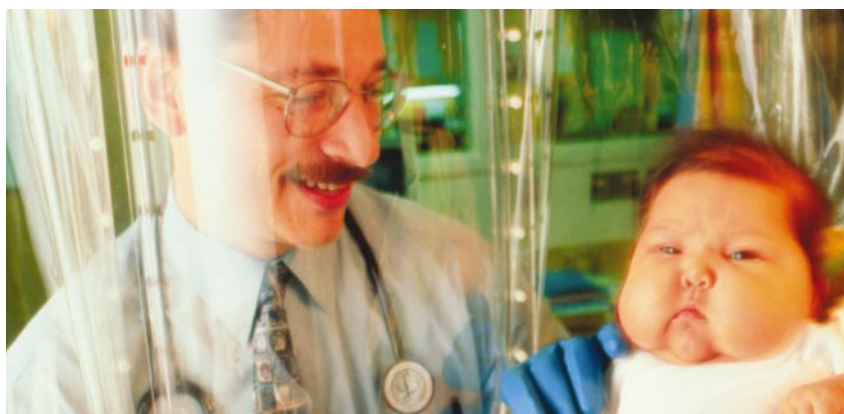
Erika Check, Washington

The revelation last week that experimental gene therapy may have given a child a cancer-like illness has provoked diverse regulatory responses around the world.

The child was involved in a trial to treat severe combined immunodeficiency disease (SCID), a condition that disrupts the early development of the immune system.

France and the United States immediately suspended all SCID gene-therapy trials, whereas Britain elected to continue enrolling and treating patients in its two studies. Germany, by contrast, had already suspended a wider range of 13 similar gene-therapy trials in June, following reports that the technique may cause cancer in mice.

The SCID patient is a three-year-old French boy taking part in a clinical trial run by gene-therapy pioneer Alain Fischer of the Necker Hospital for Sick Children in Paris. SCID is fatal if not treated by a bone-marrow transplant. But many children lack a perfectly matched donor, and of these about 30–40%



Don Kohn has successfully used gene therapy to treat children with severe combined immunodeficiency.

who suffer from a particular form of SCID die.

Fischer's technique hinges on a gene for a receptor that triggers the normal development of cells in the immune system. He uses a retrovirus to ferry the corrective gene to bone-marrow cells, and has so far cured 10 children.

But routine checks revealed worrying signs that the boy, one of those 10, had developed a leukaemia-like illness. The incident could have wide implications for the troubled field of gene therapy, as Fischer's SCID treatment had been the field's greatest success to date.

The widely diverging reactions to the news highlight the difficulty of balancing the immediate risks posed by some clinical trials against their long-term potential benefits. In the United States, the Institute of Medicine has just proposed reforms in a bid to get the balance right (see page 546). And the US Food and Drug Administration (FDA) has called a meeting this week to discuss how to proceed with gene therapies for SCID and other diseases.

Clinical researchers attribute the varying responses to the French boy's condition to a multitude of cultural, political and scientific factors. French authorities are sensitive to the fact that the adverse event occurred there, and US regulators are acutely aware of questions that were raised after the death of 18-year-old Jesse Gelsinger in a gene-therapy trial three years ago. Germany, meanwhile, maintains a generally cautious approach to research on human subjects, whereas Britain tends to take a more permissive approach, in the hope that such research will yield major benefits.

"Different countries are taking slightly different positions — the British are being

China ponders joining fusion project

Geoff Brumfiel, Washington

Chinese officials are set to meet with negotiators from the European Union, Japan, Russia and Canada in Tokyo to discuss possible involvement in ITER — formerly the International Thermonuclear Experimental Reactor — which many believe is the next step in developing fusion power.

Officials at the Chinese ministry for science and technology have written to senior negotiators in each of the four ITER member states to express a "strong interest" in Chinese participation, says Paul Vandenplas, emeritus professor of the Royal Military Academy in Brussels, Belgium, and part of the European Union's negotiating team.

Vandenplas says that China would like to become a full partner in the US\$4-billion collaboration. This would require China to contribute about \$100 million per year in cash or equipment.

The Tokyo meeting will take place just after an official site-negotiation meeting in Rokkasho, Japan, at the end of this month. Vandenplas says that ITER's member states will provide China with information about the project and will informally discuss China's participation in negotiations.

China has a small but growing fusion programme, and officials — including the country's president, Jiang Zemin — have recently shown increasing interest in the field, says David Baldwin, head of fusion research at General Atomics in San Diego, California. "People who have visited Chinese laboratories are very impressed with the depth of knowledge," he adds.

US government officials have declared that the United States plans to rejoin ITER, which it left in 1999 (see *Nature* 417, 676; 2002) but no official announcement has been made.

very open-minded and the Germans want to block everything," says Fischer. At this point, he adds, it seems reasonable to halt gene-therapy trials that use retroviruses to target bone-marrow cells.

Retroviruses insert themselves randomly into DNA, but scientists have long worried that this activity has the potential to promote uncontrolled cell growth.

In the French child's case, Fischer says, the retroviral vector has inserted itself into a stretch of DNA that regulates the gene *LMO2*. This gene can cause leukaemia, and the boy's defective immune cells seem to be expressing *LMO2*. This raises the possibility that the retroviral vector activated the gene, leading to the boy's illness.

The suspension of German trials earlier this year followed a report that a retroviral gene therapy had caused leukaemia in mice (Z. Li *et al. Science* **296**, 497; 2002). Christopher Baum, the study's principal investigator, who is currently at the Cincinnati Children's Hospital Medical Center, says that the German authorities will let the trials continue after doctors revise their informed-consent documents to tell patients about the risk of cancer.

Baum and his co-authors suspect that a particular marker gene used in their study played a part in causing leukaemia. Therefore, Baum and Fischer say, the *Science* finding by itself did not justify halting all trials that use retroviral vectors.

"Until now, this was a perfect therapy for SCID. Now there are risks," says Don Kohn, a paediatrician at Children's Hospital of Los Angeles who was leading a gene-therapy trial for SCID. Kohn has treated four children, all of whom are still healthy, but the FDA has now suspended his trial. ■

Additional reporting by Quirin Schiermeier.

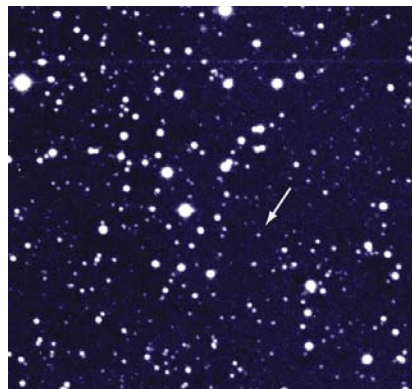
Discovery of giant asteroid gives Pluto a rocky outlook

Tony Reichhardt, Washington

Two astronomers at the California Institute of Technology have found a 1,200-kilometre-wide rock in the Kuiper Belt that circles in the region of Neptune's orbit.

The rock — named Quaoar after the creation force of the tribe that once inhabited the Los Angeles area, where the institute is based — is the largest of the 600-plus Kuiper-belt objects (KBOs) so far identified. And with a diameter half as big as Pluto's and roughly the same size as that of its moon Charon, Quaoar bolsters the argument that Pluto can also be considered a KBO.

Chad Trujillo and Mike Brown found Quaoar on a digital image taken on 4 June with the Palomar Observatory's 1.2-metre Oschin Telescope, which they set up to search



Big issue: the large size of Quaoar raises the question of whether Pluto is a Kuiper-belt object.

for objects moving against the background star field. The pace of KBO discovery has accelerated in recent years, thanks to this and other Kuiper-belt searches that can spot even fainter objects.

Since the discovery, Trujillo and Brown have used the Hubble Space Telescope, the Keck telescope in Hawaii, and the IRAM millimetre-wave telescope in Spain to collect further details about Quaoar. With the benefit of hindsight, they have identified Quaoar in astronomical images taken as far back as 1982. It appears to be spherical, or nearly so — unlike another large KBO, the 900-kilometre Varuna, which is elongated.

Quaoar is also relatively bright, reflecting 10% of the light that hits it. This adds to evidence that the 4% reflectivity assumed for KBOs may be wrong. And because the size of such objects is often inferred from their brightness, Trujillo says that many KBOs may be smaller than originally believed. Two other KBOs — Ixion and 2002 AW197 — were once thought to be as big as Quaoar, assuming 4% reflectivity. But Trujillo says that Quaoar's size has been confirmed by measurements from Hubble, whereas the others' haven't.

Alan Stern, a planetary scientist at Southwest Research Institute in Boulder, Colorado, and leader of the New Horizons spacecraft mission to reach Pluto by 2015, called Quaoar "a wonderful discovery". But its size record may turn out to be short-lived. Stern won't be surprised if someone eventually finds an Earth-sized object in the Kuiper belt. In fact, he adds, "I'd be surprised if we don't". ■

C. TRUJILLO & M. BROWN/CALTECH

Call for clinical-trial reform leaves critics unmoved

Kendall Powell, Washington

Medical watchdog groups in the United States say they are becoming frustrated that reforms of the clinical-research system have failed to materialize, despite repeated calls for tougher regulation.

On 3 October, the Institute of Medicine released a report calling for modest reforms of the system. The report, "Responsible Research: A Systems Approach to Protecting Research Participants", was written by a panel chaired by Daniel Federman, senior dean for alumni relations and clinical teaching at Harvard University.

The report calls for all research on human subjects to be federally regulated, and recommends the creation of an independent office to oversee such research, and a national registry to track participants

and adverse events in all clinical trials. But critics say that this is the fourth such report in five years, and that little has changed in the way that clinical research is regulated.

As the report was released, Senator Edward Kennedy (Democrat, Massachusetts) was introducing legislation to implement some of its main recommendations. But similar legislation has previously languished in the face of opposition from drug firms, medical schools and biomedical researchers.

"We've been talking about this for too long," says John Mather, chief officer for research oversight at the Department of Veterans Affairs. "Who is going to take that report off the table and do something with it?"

Critics of the existing system have nevertheless welcomed the report. It will get

"everyone on the same page to deal with problems in clinical research", says Paul Gelsinger, whose son Jesse died while participating in a 1999 gene-therapy clinical trial at the University of Pennsylvania.

But some argue that the report fails to recommend the tough new rules that they would like to see applied to clinical trials. "When things aren't mandatory, nobody pays attention, and that's not going to change the rate of violations, death or harm," says Vera Sharav, president of the Alliance for Human Research Protection.

The report was commissioned by the Department of Health and Human Services after Jesse Gelsinger's death, in response to concerns that institutional review boards do not provide adequate protection for research subjects. ■



Biotechnologists hoping to run crop trials in Switzerland will face a veritable mountain of regulation.

Swiss law imposes iron rule on biotechnology field trials

Quirin Schiermeier, Munich

The Swiss parliament last week passed a law setting some of the world's toughest restrictions on the conduct of trials of genetically modified crops.

The bill, which was passed on 2 October, allows trials of transgenic crops to be carried out, but only if they can jump through a set of strict regulatory hoops. It was nonetheless greeted as a qualified success by Switzerland's large agricultural biotechnology industry, which hopes that the law will end a *de facto* moratorium on such trials. But the legislation must still be debated in the Swiss second chamber and could yet be challenged in a referendum.

The new law, known as Gen-Lex, was conceived by the Swiss government to draw the sting from environmentalists' attempts in a 1998 referendum to outlaw all such trials. The referendum, which failed by a narrow majority, would have banned field trials of genetically modified crops, as well as the creation of transgenic animals (see *Nature* 393, 507; 1998).

If confirmed, the new law will require applicants to prove that their research cannot be carried out under laboratory conditions and that, if released, the crops are unlikely to cross-pollinate non-transgenic plants or otherwise contaminate the environment. All trials must also include a research component on the biosafety of the experiments.

Almost 200 researchers had signed a plea asking for a more research-friendly law. But

despite the tight restrictions, they will be relieved to at last have a realistic prospect of carrying out field trials. The lack of trials had threatened to cut off Swiss researchers from new developments in molecular plant research and botany, they say.

"The bill sets very high hurdles for the approval of transgenic plant releases, but it could have been worse," says Daniel Schümperli of the Institute of Cell Biology at the University of Bern, and chair of the Swiss Academy of Sciences' forum on genetic engineering. A step-by-step approach, he says, similar to the phases of clinical trials, should allow plant scientists to meet the requirements and minimize the environmental risks associated with field trials of transgenic crops.

The biotechnology industry also welcomed the bill's passage. "The small Swiss market is not our main concern," explains Arthur Einsele, a spokesman for Syngenta, one of the world's largest agricultural biotechnology companies, whose headquarters are in Basel. "But a moratorium would have been a very bad signal for the European Union. We have an interest in our headquarters being based in a country with a positive research climate."

But Switzerland remains deeply divided over the issue of agricultural biotechnology. The law could be modified by the second chamber of parliament, and environmental groups have said they will attempt to call a referendum if any more concessions are made to facilitate field trials.

Winning universities set out to fulfil Japan's plans for excellence

David Cyranoski, Tokyo

Japan's education ministry has named the first 113 centres of excellence to be set up at the nation's universities.

The programme, which aims to make Japanese research more competitive, saw 464 groups from 163 public and private universities compete for the awards.

"Until now, Japan has favoured stability in funding, but now it is leaning towards competition," says Yuichiro Anzai, president of Keio University in Tokyo and a co-chairman of the programme.

The winning centres will each receive grants worth between ¥100 million and ¥500 million (US\$800,000 to \$4 million) per year for a renewable five-year term — although an unfavourable review after just two years can close any of the centres.

Teams of about 20 professors competed for the grants in five fields — life sciences, chemistry and materials, information technology and electronics, humanities, and a broad category of interdisciplinary topics, including environmental science and energy research.

The chosen centres span 50 universities, with 49 of the centres based at the 7 former imperial universities. The University of Tokyo and Kyoto University lead the pack with 11 centres apiece.

But critics complain that the selection standards were not made clear — and that the government has not provided any feedback on how projects were chosen.

"Most people do not believe this programme has been considered carefully," claims one referee for the programme, who declined to be named. Some researchers also say that the money could have been better spent on fewer, larger projects.

But most scientists are optimistic that the programme will boost high-quality, multidisciplinary research. "More competition will have a great effect on Japanese universities," says Tadamitsu Kishimoto, president of Osaka University. ■



Worm cast in starring role for Nobel prize

Erika Check, Washington

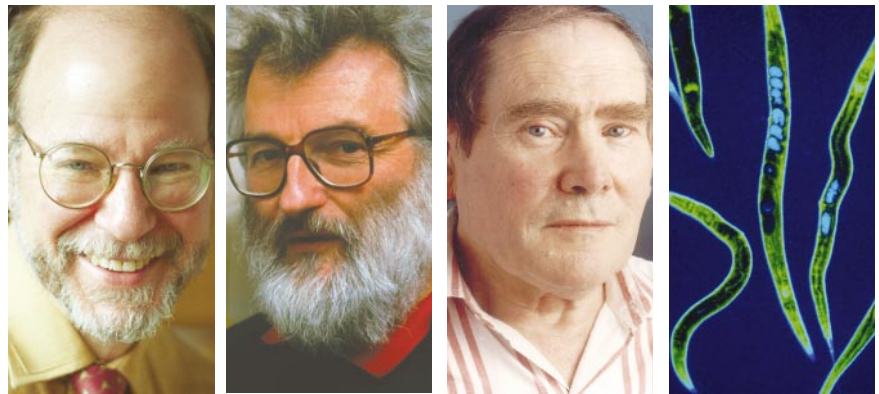
A humble nematode has wormed its way into the affections of the scientific community and helped to secure this year's Nobel Prize in Physiology or Medicine.

The award goes to three biologists whose work on the model organism *Caenorhabditis elegans* has yielded insights and spin-offs in such diverse fields as cancer research and modern genomics.

South African-born molecular biologist Sydney Brenner, president of the Molecular Sciences Institute in Berkeley, California, will share the US\$1.1 million prize with John Sulston of the Sanger Institute in Cambridge, UK, and American Robert Horvitz of the Massachusetts Institute of Technology.

"This award is fantastic and well-deserved," says Robert Waterston, director of the Genome Sequencing Center at Washington University in St Louis. "It's not only a tribute to the investigators, but also to the power of the worm."

Brenner started the field of *C. elegans* biology in the 1960s. He had already made a name for himself as a founding father of molecular biology and wanted to concentrate on the development of the nervous system. So he began looking for a model system that was



Developing a theme: (from left) Robert Horvitz, John Sulston and Sydney Brenner have won the 2002 physiology Nobel for their work on the biology of the nematode *Caenorhabditis elegans*.

simpler than the familiar fruitfly, but more complex than bacteria. He chose the nematode because of its short life cycle, genetic simplicity and tiny size—less than 1 mm long.

"We wanted to find a good experimental organism that people could study, and it has proved to be the right one," Brenner says. "What gives me the greatest pleasure is that so many people are working on this system now. When I meet scientists who are my 'great-great-grandchildren', it's very satisfying."

Sulston was one of Brenner's earliest students. During the early 1970s, he undertook the 'lineage project', in which he spent hours each day staring through a microscope at dividing *C. elegans* cells. Finally, in 1977, he published a complete map of how one fertilized egg gave rise to the 959 cells of the adult worm (J. E. Sulston and H. R. Horvitz *Dev. Biol.* **56**, 110–156; 1977).

Along the way, Sulston made several crucial observations, including the recognition

Universal view nets award for cosmic sleuths

Philip Ball, London

Three pioneers who together opened a new window on the Universe have this year been rewarded with the Nobel physics prize.

Raymond Davis of the University of Pennsylvania in Philadelphia and Masatoshi Koshiba of the University of Tokyo share half the prize for developing techniques to detect the dauntingly elusive subatomic particles called cosmic neutrinos.

The other half of the award goes to Riccardo Giacconi for his role in building the first instrument for detecting X-rays from cosmic sources. Giacconi is now president of Associated Universities Incorporated, a Washington-based consortium that runs the National Radio Astronomy Observatory for the US National Science Foundation.

Davis set up an audacious experiment in the 1960s to detect neutrinos coming from the Sun. Solar theories had predicted that nuclear fusion reactions at the Sun's heart would generate vast numbers of neutrinos. But the particles rarely interact with matter, so it was unclear whether they could be detected.

To catch the solar neutrinos, Davis directed the construction of a detector at the Homestake gold mine in South Dakota, where a tank was filled with 615 tonnes of the

cleaning fluid tetrachloroethylene. He calculated that a tiny proportion of the neutrinos would have enough energy to collide with chlorine atoms in the fluid, creating argon atoms in the process. About 20 argon atoms were expected to be created every month, and Davis devised a method of extracting them by pumping helium gas through the liquid. By the time it closed in 1994, the experiment had detected about 2,000 argon atoms.

Koshiba followed up this work in the 1980s, leading a team that built a neutrino detector known as Kamiokande. Housed in a mine roughly 240 kilometres northwest of Tokyo, Kamiokande monitored the light produced by neutrinos when they collided with water molecules, allowing it to identify when the particles arrived and which direction they were moving in. Koshiba's team showed conclusively that the Sun emits neutrinos.

In February 1987, the group also detected



Out of this world: winners of this year's physics Nobel, (from left) Raymond Davis, Masatoshi Koshiba and Riccardo Giacconi.

a neutrino burst from a supernova explosion in a neighbouring galaxy, showing that neutrino detectors could be used to study distant astronomical objects.

Koshiba's award is the third to go to Japan in as many years, keeping the country on track for the government's stated aim, set out in 2000, to try to win 30 Nobel prizes within 50 years. "My dream now is that one of my students can also get a Nobel prize," says Koshiba.

Giacconi began working on X-ray astron-

that cells die in a repetitive pattern during normal animal development. He also discovered the first gene mutation that disrupted this process of 'programmed cell death'.

"I just loved watching the cells. It's a beautiful thing to do and a challenge in the jigsaw-puzzling sense to get it all," he says.

Horvitz picked up from Sulston's work on programmed cell death in the mid-1970s. In 1986, he published a description of the first genes known to cause cell death — *ced-3* and *ced-4* (H. M. Ellis and H. R. Horvitz *Cell* **44**, 817–829; 1986). He then proved that humans have a gene similar to *ced-3*, and showed that programmed cell death occurs through similar pathways in diverse life-forms, including humans.

Some scientists also see a larger significance to this year's Nobel, especially when combined with the 1995 Nobel for work in the fruitfly and the 2001 prize, which rewarded groundbreaking work in yeast.

"These awards are a recognition that you can make major advances in medicine by studying genetically tractable model organisms," says Gerald Rubin, a vice-president at the Howard Hughes Medical Institute in Chevy Chase, Maryland. ■

Additional reporting by Declan Butler.

omy in the late 1950s and helped to launch a field that has transformed our view of the Universe. Previously, astronomers had relied on optical, infrared and radio frequencies, but X-rays revealed violent astrophysical processes that no one had expected. "X-rays show us the extreme end of things," says Martin Ward of the University of Leicester, UK. Giacconi's award is "very well deserved for half a century of work in this area," he adds.

X-rays are absorbed by the Earth's atmosphere, so Giacconi's team developed instruments to detect them from rockets. Their initial aim was to see whether solar radiation caused X-ray emission from the Moon, but once beyond the atmosphere, their detector revealed other X-ray sources such as that in the constellation of Scorpius.

Strikingly, all three prizewinners have been recognized largely for their work on instrumentation — an acknowledgement, perhaps, that many questions in high-energy astrophysics cannot even be framed until observational methods exist. ■

Additional reporting by David Cyranoski.

The Nobel chemistry prize was announced after *Nature* went to press. For coverage see

♦ www.nature.com/nature



Beer with me: Arnd Leike tries not to slur his words as he accepts his Ig Nobel prize in physics.

Bubble won't burst for spoof Nobels

Steve Nadis, Boston

The conversation last Thursday in a pub near Harvard University was about beer — but the discourse was perhaps a tad more elevated than the usual drunken banter.

For his work on the exponential decay of beer froth, Munich University physicist Arnd Leike was due to receive the 2002 Ig Nobel prize in physics. His companion that evening was Harvard chemist Dudley Herschbach, a Nobel laureate.

Herschbach pointed out that small beer bubbles disappear first, in the same way that small black holes evaporate faster than big ones. "There's something cosmic going on here," he said. The two then strolled to the nearby Sanders Theatre for the twelfth annual awards ceremony of the Ig Nobels — Harvard's enduringly popular spoof on the real Nobel prize awards.

"Every winner was chosen for work that first makes people laugh and then makes people think," explains the event's organizer, Marc Abrahams.

This year's ceremony featured the world premiere of *The Jargon Opera*, 'a jargon-free mini-opera in four acts', gamely featuring Nobel laureates Herschbach, William Lipscomb and Richard Roberts.

David King, chief scientific adviser to the UK government, was also there. His appearance may have healed a rift that opened in 1996 when his predecessor Bob May warned that the Igs risked damaging science by ridiculing and trivializing worthy work (see *Nature* 383, 291; 1996). "I don't want to be critical of Bob, but I think this is all in good fun," King says.

King was personally able to congratulate his countryman Chris McManus of

University College London, who captured the medicine prize for a 1976 *Nature* paper, "Scrotal asymmetry in man and in ancient sculpture".

Navel-gazing brought a prize for Karl Kruszelnicki of the University of Sydney. His survey of the qualities of human belly-button fluff earned him the interdisciplinary research prize. And Charles Paxton of the University of St Andrews, UK, accepted the biology prize for a report he co-authored claiming to have found that sexual arousal in ostriches is enhanced by the presence of humans.

Vicki Silvers of the University of Nevada-Reno and David Kreiner of Central Missouri State University shared the literature prize for their study, "The effects of pre-existing inappropriate highlighting on reading comprehension". The message of their paper, said Kreiner, is "don't buy a textbook that has been highlighted by an idiot".

Theo Gray of Wolfram Research won the chemistry prize for creating a three-dimensional periodic table of the elements that also serves as a coffee table. During his acceptance speech, Gray apologized to his parents for dropping out of graduate school, noting that "this is as close to academic distinction as you're going to get out of me".

Leike toasted the audience and sipped beer from a graduated cylinder. "I hope my paper will help get people interested in science," he later told *Nature*.

Herschbach agrees with the sentiment. "The whole reason for taking part in something like the Igs is to convey the message that scientists are really kids at play," he says. ■

Haven will be first glimpse of jungle for retired chimps

San Diego Up to 200 chimpanzees owned by the US National Institutes of Health (NIH) have gained a retirement home. The NIH last week awarded Chimp Haven — a non-profit organization formed by researchers, animal-welfare advocates and concerned citizens — \$24 million to build and run a jungle-like compound on 200 acres of donated land near Shreveport, Louisiana.

Most of the animals were born in captivity and have never lived in the wild. Some have been used in research, including studies of HIV and hepatitis. Formerly wild chimps will be used to teach the lab animals foraging methods in a natural setting. “They have never climbed a tree, made a nest or fished for termites,” says Linda Brent, an anthropologist at the Southwest Foundation for Biomedical Research in San Antonio, Texas, and president of Chimp Haven.

The announcement comes after animal-rights activists and a charitable foundation last month purchased the troubled White Sands primate-research facility in New Mexico with the aim of turning it into a private retreat (see *Nature* **419**, 330; 2002).

► <http://chimplhaven.org>

European gamma-ray satellite ready for launch

Washington A satellite designed to study black holes, supernovae and high-energy flashes of radiation known as gamma-ray bursts is gearing up for launch.

The European Space Agency’s International Gamma-Ray Astrophysics Laboratory (INTEGRAL) is due to lift off from Baikonur in Kazakhstan on 17 October.

A successor to NASA’s Compton Gamma-Ray Telescope, which was taken out of operation in 2000, INTEGRAL will train higher-resolution instruments on selected targets from an elliptical orbit that will stretch 150,000 km from Earth. This will place it above Earth’s radiation belts — regions of charged particles that have been trapped by Earth’s magnetic field.

Along with its γ -ray detector and



Getting the green light: INTEGRAL undergoes final preparations before next week’s launch.

spectrometer, the observatory will carry instruments that can simultaneously monitor X-rays and visible light from the same object. Built to last for at least two years, INTEGRAL has cost the European Space Agency 330 million euros (US\$320 million) — not including launch costs, which Russia will meet in exchange for observing time.

AIDS activist freed by Chinese authorities

Tokyo Wan Yanhai, the AIDS activist who was arrested in China this summer, has been released after admitting to leaking Chinese state secrets.

Wan, a consistent critic of the Chinese government’s handling of the country’s AIDS problem, is well known among human-rights activists for promoting AIDS awareness via his website. He was arrested on 24 August after posting data on the Internet about the spread of AIDS in certain Chinese villages. The government claimed that some of the information he made public was classified (see *Nature* **419**, 99; 2002).

Wan was released last month after admitting to having disclosed classified data. It is not yet known what penalty he will face.

The New York-based organization Human Rights Watch is one of several groups that have since expressed concern that, despite his release, Wan will be effectively silenced and will be unable to continue his AIDS educational activities.

\$2,000 reward to sharpen focus on asteroids

Washington Short of cash? A telescope could provide the answer. The US House of Representatives last week passed a bill that will grant \$2,000 prizes to amateur astronomers who discover asteroids that cross Earth’s orbit.

The awards, proposed by House Space and Aeronautics Subcommittee chairman Dana Rohrabacher (Republican, California), would go to US citizens who discover bright Near-Earth Objects or who help to pinpoint and catalogue known asteroids. Rohrabacher hopes that the Senate will pass a similar bill to get the programme under way.

Meanwhile, a government-funded asteroid search is progressing well. NASA has already catalogued 619 of the estimated 1,000 asteroids in Earth’s vicinity that are more than one kilometre across, says the agency’s science chief Edward Weiler, and is on track to identify 90% of them by 2008.

But Weiler says he is not keen for NASA to tackle the next logical mission — tracking down the 50,000 or so objects between 300 metres and one kilometre in size. He suggests that the search is better left to an agency such as the National Science Foundation, which funds ground-based astronomy.

'GI bill' proposed to target postgraduates

Washington A system of fellowships is needed to revive interest in postgraduate science and technology among US students, a President's Council of Advisors on Science and Technology panel said last week.

Wayne Clough, president of the Georgia Institute of Technology and head of the panel, said that half of all US engineering doctorates go to foreign students. Only 6% of US undergraduate degrees go to science or engineering students, compared with 10% in the United Kingdom and 8% in Japan. Engineering, science and technology workforces will suffer unless the numbers improve, the panel said.

Members envisage a nationally funded graduate fellowship that would provide three or four years of support. This would give graduate students more security than relying on year-to-year assistantships, says Clough. He suggests that the programme might work like the popular GI bill, which pays for the education of military personnel but requires some postgraduate service to the agency that provided the fellowship.

Iguana expert wins payout after oil spill wrecks study

London An ecologist whose studies of marine iguanas in the Galapagos Islands were ruined by an oil spill is to receive compensation as part of a US\$10-million damages package awarded to Galapagos National Park.

Princeton University researcher Martin Wikelski made a US\$600,000 claim as part of a wider action brought by the park against the insurers of the oil tanker *Jessica*, which ran aground in January 2001 (see *Nature* 417, 578; 2002). A court in Ecuador found in favour of the park and against British insurance company Terra Nova on 3 October.

Wikelski, who has been studying the iguanas since 1987, found that more than 60% of them died at sites affected by the oil. He says that he will use the money to set up a trust fund for young Ecuadorian scientists. Terra Nova had been given three days to appeal, but a spokesperson for the company declined to comment on the company's decision as *Nature* went to press.

Burials reveal human sacrifice in Peru

Lima Peruvian archaeologists have found more than 200 bodies, believed to have been those of people killed in a sacrificial ritual, near the coastal town of Casma.

The bodies are thought to date from about 1350, nearly 200 years before Spanish colonialists arrived. The ritual is believed to have been carried out by the Chimú people who lived in Peru at the time. The victims had been bound, stabbed through the heart,



Grave news: discovery of a massacre proves that Peruvians practised large-scale human sacrifice.

then buried facing the sea. Archaeologists believe that they were probably fishermen because nets and other fishing equipment were found in some of the graves.

The site, which was discovered in 1998 but announced only recently, is the largest of its kind found in South America, and represents the first evidence that the Chimú people carried out human sacrifices.

Controversial minister loses job in Japanese reshuffle

Tokyo Three science-related ministers lost their positions in last week's Japanese cabinet reshuffle.

Koji Omi, the minister for science and technology, ended a 17-month term that had been dominated by his plans for industry–university collaborations. Some initiatives ruffled the feathers of researchers, who felt that Omi placed too much emphasis on research commercialization (see *Nature* 412, 364; 2001). He has been replaced by Hiroyuki Hosoda, from the economics ministry.

Another casualty, agriculture minister Tsutomu Takebe, recently stated his desire to remain in the position despite criticism of his ministry's handling of Japan's mad cow disease outbreak. "If everyone retires when things go wrong, nothing will ever get fixed. I want to stay around and fix this problem," Takebe said in July. In his year in the job, Takebe began to overhaul the ministry and helped to establish a new food-safety agency (see *Nature* 418, 357; 2002). Takebe has been replaced by Tadamori Ooshima.

The third post, that of environment minister Hiroshi Oki, has gone to Shunichi Suzuki, who — like Ooshima — is a member of the ruling Liberal Democratic Party.

Wired for success

Nanowires, nanorods or nanowhiskers. It doesn't matter what you call them, they're the hottest property in nanotechnology. David Appell investigates.

In the much-hyped world of nanotechnology, the carbon nanotube has hogged the limelight. These tiny hollow tubes, which resemble rolls of wire mesh, have already been fashioned into miniature tweezers that could one day be used to pick up and move individual cells. They might also form the basis of nanometre-sized computer circuits, according to advocates¹. And some NASA researchers have even suggested that the tubes could be used to build space elevators to ferry people and satellites into orbit.

But away from the spotlight, nanotube researchers have hit problems. Potential computing applications have generated the most excitement, but some of the tubes' crucial electronic properties, which depend on precise orientation of the carbon mesh, are hard to control. The difficulties may not be insurmountable, but they have persuaded some scientists to turn their attention elsewhere. Over the past few years, alternative nanoscale structures have attracted attention. One in particular — called either the nanowire, nanorod or nanowhisker, depending on which researcher you ask — has gained an impressive following.

These nanometre-diameter wires, which can be grown from a variety of different

materials and can reach tens of micrometres in length, have one major advantage — unlike nanotubes, their chemistry is relatively easy to tailor. By growing nanowires from different semiconductors — materials whose conductivity lies somewhere between those of a conductor and an insulator — some researchers have already used them to build transistors and diodes. Others have constructed lasers and biological sensors. "There are a lot of unknowns to be discovered," says Charles Lieber, a Harvard University chemist who has abandoned much of his work with nanotubes to focus on nanowires. "If you're creative, there are very different things to do with these materials."

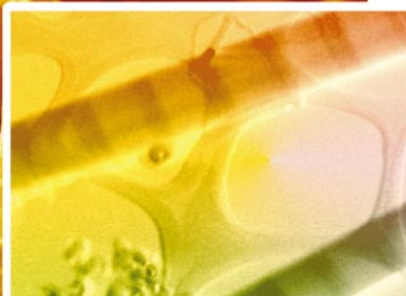
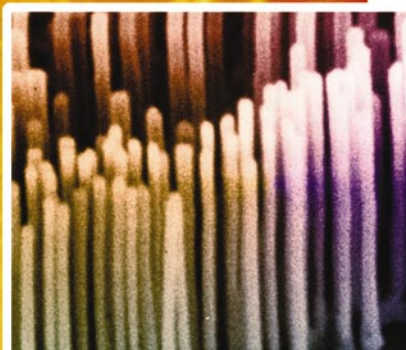
The chips are down

Both nanotubes and nanowires are attracting attention because of a looming crisis in computing. Traditional electronic circuits are built by etching individual components into silicon wafers. Advances in this technique have ensured that computer chips have steadily shrunk in size over the past few decades. But the circuits can only become so small — current fabrication methods will fail well before the size of electronic components reaches atomic dimensions, probably within the next 20 years.

Ahead by a whisker: semiconductor nanowires (main picture) are relatively easy to make. They can be aligned in channels (inset, top), grown in arrays (inset, middle) or produced from alternating materials as superlattices (bottom).

Electronic circuits built from nanowires or nanotubes are one way in which the miniaturization trend could be continued. The wires on today's cutting-edge computer chips are a few hundred nanometres thick — nanowire or nanotube alternatives could be at least an order of magnitude thinner. Other more complex components of electronic circuits, such as switches and light-emitting diodes, could also be built from nanowires or nanotubes.

Progress with nanotubes has been impressive since they were first discovered just over a decade ago². For example, transistors — electronic devices that can act as switches and can amplify current — have been built from nanotubes, and wired together to form logic gates, the building blocks of computers¹. But to create these devices, researchers must carefully tweak the electronic properties of the nanotubes. This is not always easy to do, as these properties depend on the diameter and crystal structure of the tubes — parameters that are difficult



to manipulate. "Carbon nanotubes have unique attributes," says Lieber. "But one has a difficult time controlling their metallic and semiconductor properties."

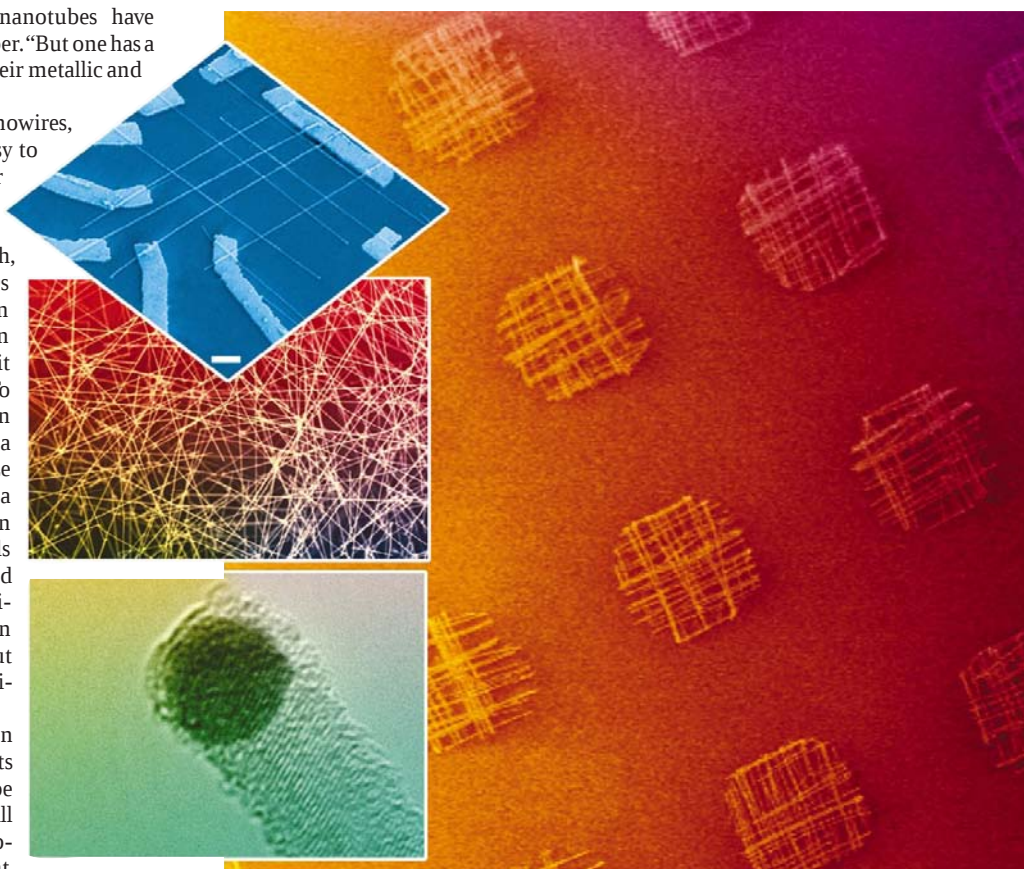
The composition of nanowires, by contrast, is relatively easy to control. The most popular method for producing them involves vapour-liquid-solid (VLS) growth, first proposed in the 1960s but only coming into its own for nanowires in 1998, when Lieber's group combined it with other techniques^{3,4}. To use VLS to make a silicon nanowire, for example, a laser is first used to vaporize the silicon together with a metal catalyst, typically iron or gold. As the vapour cools it forms nanosized liquid clusters containing both silicon and some iron. Silicon wires then slowly grow out from the cluster, as more silicon atoms condense.

If the nanowire is grown from a semiconductor, its electrical properties can be altered by adding small amounts of a second substance, called a dopant. Adding boron to a silicon nanowire, for example, increases the number of 'holes' within the network of electrons available for conduction, whereas adding phosphorus creates additional free electrons⁵. James Ellenbogen, a nanowire researcher at MITRE, an engineering and computing firm in Bedford, Massachusetts, says that this ability to control the composition of nanowires is key. "One word distinguishes what semiconductor nanowires have that carbon nanotubes don't," he says. "Chemistry."

Up the junction

In conventional circuits, combinations of differently doped semiconductors are used to build electronic devices, and the same can be done using doped nanowires. Last year, for example, Harvard chemist Yi Cui, together with Lieber, created two differently doped silicon semiconductor nanowires and joined them together. The difference in the density of conducting electrons has a very useful effect on the junction where the wires meet: it behaves as a diode, allowing current to flow in only one direction across it⁶.

This January, Yu Huang, another member of Lieber's group, described how he formed a cross using two different nanowires — one made of gallium nitride, the other of silicon — to create a transistor⁷. Such transistors could be the basis of a nanowire memory unit. The Harvard team has also coated vari-



Memories are made of this: crossed nanowire arrays (main picture) and an experimental 4 × 4 array (inset, top) are being tested for use as computer memory. The nanowires (inset, middle) are readily grown on metal catalysts such as a gold nanoparticle (black spot, inset, bottom).

ous nanowires with charged molecules and shown that the charge affects the nanowires' ability to conduct electricity. By applying a voltage to change the charge, the researchers could switch the wire's conductance between two different states⁸, in effect creating a nanowire memory that stores information in the form of the wire's conductance.

Speaking at the Nanotube 2002 conference, held this July in Boston, Lieber described how he had arranged four nanowires in the shape of a noughts and crosses (tic-tac-toe) grid. Each of the four junctions where two wires cross acts as a transistor. The conductance of the junction can be shifted between two states by applying a voltage across the two wires in the junction, allowing the system to act as a 4-bit memory.

Such a device could help the Molecular Electronics (Moletronics) programme run by the Defense Advanced Research Projects Agency, part of the US Department of Defense, to meet its goal. By September 2004, the Moletronics researchers hope to create 16 kilobits of self-assembled nanocomputer memory, possibly using nanowires, and to be able to interface the memory with macroscopic devices. Researchers at MITRE and other Moletronics-funded groups have

begun designing and simulating architectures to scale up Lieber's array. They aim to create a system of 125 × 125 nanowires, about 5 micrometres long and only a few nanometres thick — if they succeed, this would come close to providing the 16 kilobits of memory.

Nanowire memories could also be far cheaper to run than existing ones. Conventional transistors need to be refreshed up to 1,000 times a second, or the charge on the transistor leaks away and the information is lost. But the charge on Lieber's nanowires remains in place for about 20 minutes⁸. If such advantages could be scaled up, power demands for such devices could be cut by a factor of a million.

Battle of the bands

Since the start of this year, three groups — Lieber's⁹ and those of Peidong Yang at the University of California, Berkeley¹⁰, and Lars Samuelson at Lund University in Sweden¹¹ — have described how to build more complex functions into a single nanowire. By alternating between two vaporized materials during VLS, they have grown striped or 'superlattice' nanowires consisting of up to 21 alternating bands.

Each stripe contributes its own electronic

properties to the nanowire, with some useful results. A nanowire consisting of alternating bands of differently doped indium phosphide, for example, glows when a current is passed through it and could form the basis of a nanowire light-emitting diode⁹. Yang has suggested that the thermal properties of some striped nanowires could make them suitable for building switches that respond to changes in temperature¹⁰. Moreover, the technique used to make superlatticed nanowires is cheap and efficient — millions can be created in just one hour at relatively low cost. “They have opened up a new world for us,” says Lieber.

So can these individual components be assembled into a working device? The backers of the Moletronics programme clearly think so. But for practical applications, researchers will have to learn how to manipulate nanowires — a tougher challenge than one might expect. The random movement of electrons in nanowires creates areas of positive and negative charge, which causes two wires to cling together. “Some people would think it would be very hard to get nanowires to stick together,” says Ellenbogen. “Actually it’s very hard to pry them apart.”

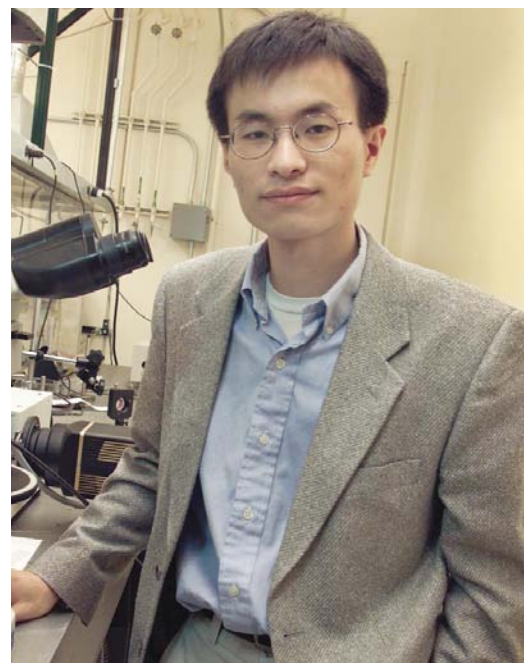
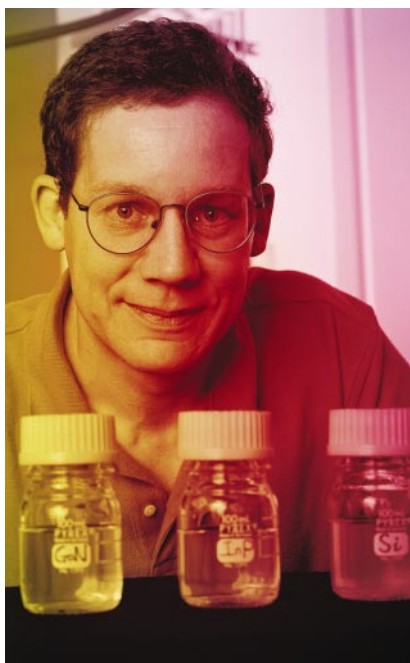
This sticky problem can be avoided by manipulating the wires in solution. Long organic molecules are added to the outside of the nanowires to act as a physical barrier. Yang’s group has manipulated nanowires by filling a trough with a nanowire solution and increasing the pressure on the liquid. This causes the nanowires to condense out of solution and align themselves side by side¹².

Crossed arrays can be formed by rotating the apparatus 90° and repeating the process. By combining layers with different electrical properties, the junctions where the wires cross can be made to function as diodes or transistors, although all of the arrays built so far are proofs of principle rather than functioning devices.

Blazing a trail

In the meantime, other potential applications are being explored. Last year, for example, Yang’s group created nanowires that emit laser light of an ultraviolet¹³ or deep blue colour. The team also showed that zinc oxide nanowires can be used to guide light, in much the same way as an optical fibre¹³. These advances lend hope to the idea that electronic circuits built from nanowires could eventually be integrated with optical systems, such as the fibres that carry voice and Internet data over long distances.

Nanowires might also form the basis of high-definition televisions. The electrons needed for the cathode-ray guns within television screens can be pulled from the tips of either nanowires or nanotubes, and the sharp ends of these structures allow lower-level electric fields to be used. Elec-



Keen edge: Charles Lieber (left) and Peidong Yang have been impressed by the versatility of nanowires.

tronics company Samsung has recently produced a prototype display using carbon nanotubes. Yang envisions a display that uses a single nanowire as one pixel, giving 10¹¹ pixels per square centimetre — orders of magnitude beyond what can be achieved today.

New chemical and biological sensors could also be built from nanowires. Many molecules change the conductance of a nanowire when they bind to its surface. By monitoring this property, Lieber’s group has built a nanowire sensor that can measure the pH of a solution¹⁴. And in unpublished work, the team has found that by placing antibodies along the wire, it can serve as a sensor for prostate-specific antigen (PSA), which marks the presence or recurrence of prostate cancer. The device can detect levels of PSA four times smaller than is now possible with blood tests that can take days, says Lieber.

The PSA sensor has been licensed to Nanosys of Palo Alto, California, a company set up by Lieber in 2001 that has already raised more than \$15 million in two rounds of funding. Such devices look promising, but whether Nanosys or another company will ever achieve nanotechnology’s main goal — to compete with conventional computers — remains unclear.

Working nanowire memory units would be tiny and cheap to run, but they will not become part of everyday computers in the near future. Researchers still have to figure out how such delicate components can be connected to the relatively gigantic wires in conventional microchips. They will also have to show that such units can compete with traditional memory systems in terms of reli-

ability — a tough challenge given the sums that have been invested in developing conventional computers.

Lieber is unsure when nanowires will replace traditional computers — if at all. “Can we actually ever exceed what’s been done? I think that it’s too early to say,” he says. “There’s a big jump between the promise and the real technology. It will come, but people need to be a little patient.”

But with nanowires showing so much potential, both researchers and funding organizations should be happy with the pace. Applications have mushroomed in the few years since VLS techniques were introduced, and many new directions await exploration. At the very least, nanowires look set to share the nanotechnology spotlight over the next few years. “Some of the things that have happened have exceeded anybody’s expectations,” Ellenbogen says. “I think there are more surprises along the way.” ■

David Appell is a freelance science writer in Gifford, New Hampshire.

1. Ball, P. *Nature* **414**, 142–144 (2001).
2. Iijima, S. *Nature* **354**, 56–58 (1991).
3. Morales, A. M. & Lieber, C. M. *Science* **279**, 208–211 (1998).
4. Duan, X. & Lieber, C. M. *Adv. Mater.* **12**, 298–302 (2000).
5. Cui, Y., Duan, X., Hu, J. & Lieber, C. M. *J. Phys. Chem. B* **104**, 5213–5216 (2000).
6. Cui, Y. & Lieber, C. M. *Science* **291**, 851–853 (2001).
7. Huang, Y., Duan, X., Cui, Y. & Lieber, C. M. *Nano Lett.* **2**, 101–104 (2002).
8. Duan, X., Huang, Y. & Lieber, C. M. *Nano Lett.* **2**, 487–490 (2002).
9. Gudiksen, M. S., Lauhon, L. J., Wang, J., Smith, D. C. & Lieber, C. M. *Nature* **415**, 617–620 (2002).
10. Wu, Y., Fan, R. & Yang, P. *Nano Lett.* **2**, 83–86 (2002).
11. Björk, M. T. et al. *Nano Lett.* **2**, 87–89 (2002).
12. Kim, F., Kwan, S., Akana, J. & Yang, P. *J. Am. Chem. Soc.* **123**, 4360–4361 (2001).
13. Huang, M. H. et al. *Science* **292**, 1897–1899 (2001).
14. Cui, Y., Wei, Q., Park, H. & Lieber, C. M. *Science* **293**, 1289–1292 (2001).

A temple of knowledge

The Library of Alexandria was the ancient world's premier seat of learning — its eventual destruction an intellectual tragedy. Can its spirit be revived in modern-day Egypt? Alison Abbott visits the Bibliotheca Alexandrina.

Euclid wrote his *Elements of Geometry* there. Herophilus identified the brain, rather than the heart, as the centre of intelligence. Eratosthenes estimated the Earth's circumference with an error of just 140 kilometres. And Hipparchus calculated the year's length to within 6.5 minutes.

Founded in the fourth century BC within the Mouseion — a shrine to the Muses that also housed botanical gardens, a zoo, dissecting rooms and an observatory — the Library of Alexandria was a vast repository of knowledge for the Greek civilization that then dominated the Mediterranean. Its founding librarian, Demetrius, and his successors operated an energetic, sometimes ruthless, acquisition policy. They employed armies of scribes and teams of translators to copy manuscripts borrowed from other collections, or impounded from ships entering the city's harbour. At its peak, the library is thought to have contained some 700,000 manuscripts. But six centuries after its foundation, after being destroyed by a series of fires, the library disappeared — along with the civilization that had nurtured it.

Some 1,700 years later, the library's spirit has been invoked in a dazzling, US\$120-million building, which will be inaugurated next week. Known as the Bibliotheca Alexandrina, the complex will include museums and research institutes, a planetarium and a conference centre. Its centrepiece is a library that will eventually hold 4 million books, 100,000 manuscripts and 50,000 maps — as well as a world of digital material.

But modern Alexandria is no longer a leading intellectual centre, and Egypt's science base is rudimentary. The grandiose project has been criticized for concentrating money that some say should have been spent on improving the region's scientific infrastructure more generally. Some critics have questioned how many people will use the library. "People felt that the library was a big PR exercise," adds

Hisham Kassem, publisher of the *Cairo Times*. "There was also a feeling that it is hypocritical, to say the least, that a book may be accessible inside the library, but banned outside its doors." The *Cairo Times*, he notes, has suffered its share of Egyptian government censorship.

Such criticisms are no longer fair, retorts Ismail Serageldin, who became the Bibliotheca Alexandrina's first director in April last year. "In a time of xenophobia, fundamentalism and obsessionism, the library stands for rationality, dialogue and scientific method," he says. Its three museums, of antiquities, manuscripts and the history of science, represent the spectrum of Alexandria's cultural heritage: Pharaonic, Greek, Muslim and Christian. Most of the funds would simply never have been made available for other scientific or cultural endeavours, Serageldin argues.

The project has certainly established an impressive physical presence, 15 years after Egypt approached the Paris-based United Nations Educational, Scientific and Cultural Organization (UNESCO) for help. Egyptian funds were secured in part by the patronage of Suzanne Mubarak, the wife of the country's president. UNESCO attracted US\$65 million from a consortium of Arab countries, plus further offers of cash, expertise or books from other donors. It also

organized an architectural competition, won in 1990 by a Norwegian company.

The result is an edifice of Pharaonic proportions. The building, on the site where the original library is thought to have stood, exudes symbolism. Placed on a shoreline that is otherwise populated by shabby high-rises, it is designed as a circle, inclined towards the sea, dipping into the pool of water that surrounds it. The design represents "the Sun rising from the sea each day to greet new knowledge", a guide explains. On its south side, separated only by a raging highway, is the Mediterranean Sea. To its north, the library is sheltered by a huge, curved wall inscribed with letters and symbols from scripts of the modern and ancient worlds.

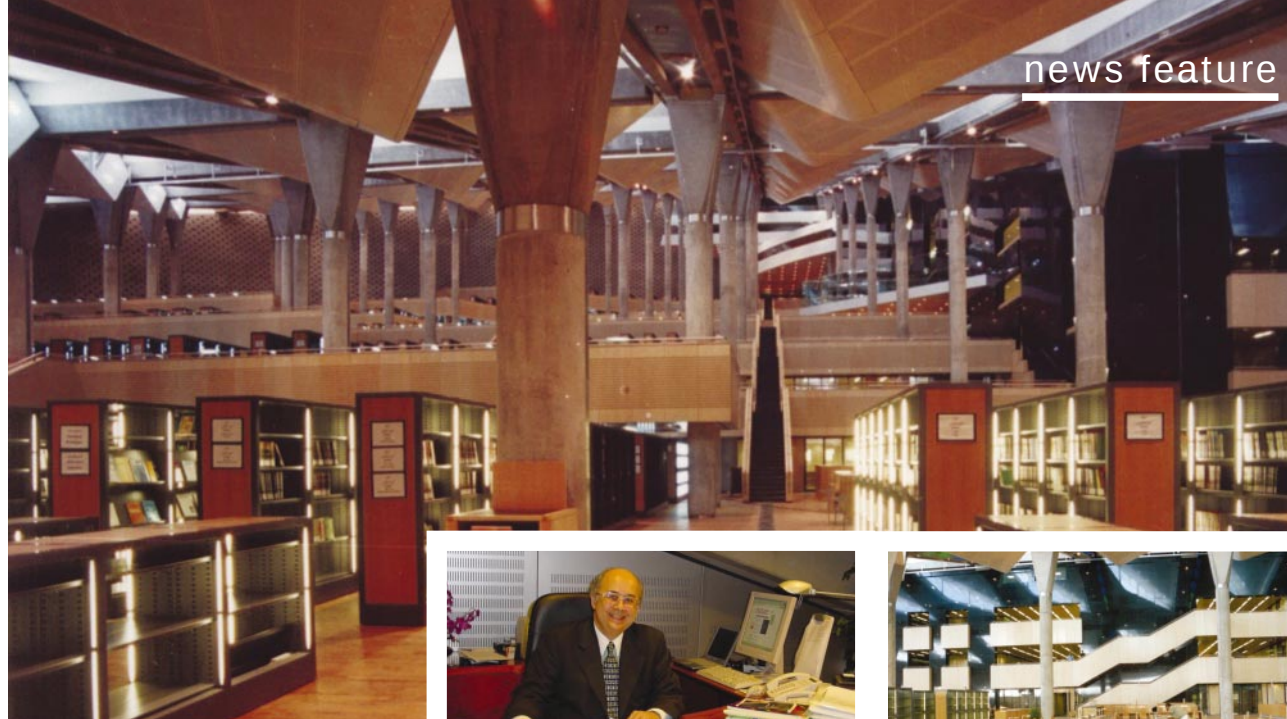
The circular roof is largely glass, allowing sunlight to illuminate the seven terraced floors of the library's reading room. The reading area is the largest in the world, with 2,000 desks, one-third of them already equipped with computer terminals. Its interior walls are faced with oxidized brass and black granite. The furniture is ergonomic and, like the building itself, is built to last.

Aware of the demise of the library's predecessor, the architects have incorporated a system of fire curtains that can descend to separate the vast open space into five sealed zones.

Walking into these surroundings, you could be in the richest, most learned country in the world. But Serageldin acknowledges that establishing a world-class library amid the poverty and bureaucracy of his native Egypt is a major challenge. "I'm always



A statue of Thoth, protector of the scribes, watches over the Bibliotheca Alexandrina's reading room.



Man with a plan: Ismail Serageldin (right) wants the library's scholarship to match its splendour.



fighting with bureaucracy," he sighs. Fund-raising will also remain a large part of his job. "We need about US\$20 million to \$25 million per year to run the library," says Serageldin. One-quarter will come from the Egyptian government, but an equal sum will have to be obtained in short-term project grants, and the remaining half from endowments.

Serageldin brings to the project the interests of a polymath and three decades of management experience with the World Bank, where he served for several years as vice-president, overseeing the bank's efforts to promote sustainable development. He has published — in Arabic, English and French — in areas as diverse as architecture and molecular biology. "I feel my whole life has been a preparation for this job," he says.

Slowly, Serageldin is bringing order to a project that had threatened to drift. Over the past 18 months he has organized the Bibliotheca's statutes and focused its aims. Gone is the high-flown idea that the library should be a universal repository of knowledge — a claim that even the US Library of Congress, with more than 18 million books, does not make. "We want to establish ourselves as an international centre of excellence in particular fields," says Serageldin. "Our themes will be science, humanities and culture; also development, where water and gender will be the main areas." And the library will have a regional focus, accumulating material on its ancient predecessor, as well as on Egypt and the Mediterranean region.

A tour of the shelves, on which 230,000 books currently rattle around, reveals that acquisitions were more eccentric before Serageldin's arrival. Alongside fascinating collections on the history of embalming and ancient medicine sit handbooks on home repair and

decorating, and titles such as *Strength Training for Football*. "People were dumping books on the library," says Mohammed Aman, dean of information studies at the University of Wisconsin-Milwaukee, who has served as a consultant on the project. "The library was in the numbers game, not the quality game, but this is clearly being reversed now."

It may take two decades for the library's shelves to be filled. But Serageldin argues that the relative paucity of the current collection should not be a subject for criticism. The Library of Congress, he points out, built its current collection from just 6,500 books purchased from Thomas Jefferson in 1815.

Serageldin also stresses that the Bibliotheca's digital arm is at least as important as its books and papers. The library has already received many gifts of digitized manuscripts, including the only papyrus that survived the destruction of the ancient Library of Alexandria. It is also digitizing its own collection of 6,700 historical manuscripts, most of which were acquired from local institutes. One of the first to receive the electronic treatment is the tenth-century Arab scholar Ibn al-Haytham's *The Trace on the Moon's Face*, a sophisticated discussion of different theories that had been propounded to explain the channels visible of the Moon's surface. Browsers in the library may scan the manuscript, its Arabic transcription, and its English, French and German translations.

The Bibliotheca also plays host to the only mirror site of the Internet Archive, a project based in San Francisco that saves copies of otherwise ephemeral web pages — you can, for example, use the archive to follow the evolution of *Nature's* homepage since 1996. And the Bibliotheca will set up the Arabic part of the Million Book Project, an international

effort funded by the US National Science Foundation to digitize seminal books in various fields and languages. However, Serageldin warns that it will take another couple of years for the library's digital infrastructure to reach full capacity.

Another task is to set up a new research institute in a discipline that has yet to be defined, which Serageldin wants to serve as a catalyst in building collaborations between Egyptian centres of excellence and developed-world institutes. Such links may be vital for the Bibliotheca's success. Foreigners are unlikely to choose to work in Alexandria — the pay is too low. But Serageldin hopes to pay enough to attract the best in Egypt, and he is banking on collaborations with the United States and Europe to maintain a high scholarly level.

In this regard, however, the current tension in the Middle East poses a threat. The library's inauguration was originally supposed to have taken place in April, but was postponed because of the escalation of the Israeli-Palestinian conflict. Next week's ceremony will proceed against the backdrop not only of these continuing events, but also of threatened US military action against Iraq. Further instability within the Arab world can only make it more difficult for Egyptians to work with American or European colleagues — and perhaps play into the hands of fundamentalist elements who are suspicious of the library's goals.

As he prepares for next week's ceremony, however, Serageldin is doing his best to put these worries to the back of his mind. "I am proud to be following in the footsteps of Demetrius," he says.

Alison Abbott is *Nature's* senior European correspondent.

► www.bibalex.gov.eg

► www.unesco.org/webworld/alexandria_new

Taxonomy needs evolution, not revolution

Some changes are clearly necessary, but science cannot be replaced by informatics.

Sir — H. C. J. Godfray in his Commentary¹ suggests that we throw out the past mechanisms of doing systematics and begin anew in a revolutionary 'brave new world' of unitary, web-based taxonomy, each group under the administration of an authoritarian body. We agree with Godfray that problems of synonymy can be frustrating and that the web is important in the dissemination of taxonomic information, but do we need a new mechanism to do taxonomy? Working within the current enabling conventions is more positive and practical than throwing them out and beginning again.

The international codes of zoological, botanical and bacteriological nomenclature are the rulebooks that have long governed how organisms are named, providing clear instructions on how to go about the process — a facility either lacking or in the process of being invented in other disciplines concerned with naming things.

At the core of the zoological and botanical codes is the type specimen. Types bring order and stability to taxonomy because they are the specimens upon which the original author based his or her descriptions, and which ultimately fix the name. Even DNA taxonomy needs archived voucher specimens to ensure that future generations can check and replicate findings. The codes operate for taxonomists as conventions, which, like UN conventions, are international in scope and have evolved during decades of intellectual debate.

Taxonomy provides a vocabulary to discuss the world: specifically, the names of organisms. We taxonomists therefore have a special role in communicating to the widest possible set of audiences. The Internet, being a distributed system of linked information, is ideal for this purpose — the various approaches under way^{2,3} will revolutionize taxonomical informatics in a positive, community-driven way.

Taxonomists have not yet made as much use of the web as they could, but first we need to identify with confidence core components of essential information, such as species lists, identification keys and illustrations. From our privileged position in the information-rich developed world, we run the risk of providing merely what we ourselves need and forgetting those in whose countries most of the diversity on Earth resides. Information is critical for global conservation, and much of it, particularly original descriptions of species

and the specimens on which these are based, remains inaccessible to those who need it most.

Building a 'biodiversity commons' (ref. 4), where access to original taxonomic descriptions is part of the distributed web system, is critical. Much taxonomic information is already available online: FishBase (www.fishbase.org) and the Cycad pages (<http://plantnet.rbgsyd.gov.au/PlantNet/cycad/index.html>), to name just a couple, are excellent examples. At present, we need meta-databases to link dispersed information, which is where the Global Biodiversity Information Facility (GBIF) is concentrating its efforts.

Web-provided taxonomy is clearly the way for the future, but the technologies needed for this to operate successfully on the scale required are only starting to be available, and quality control is something that must also be addressed⁵. The Natural History Museum is putting on its website (www.nhm.ac.uk), as a priority, information about its type specimens, which will provide the kind of focus for scientifically rigorous and accessible

taxonomy that Godfray advocates.

Informatics, however, is not a substitute for science. To ensure worldwide consistency and accuracy we also need open access to the foundations of our taxonomy, by making high-quality illustrations and descriptions of the type specimens available on the web. For an institution such as ours, housing nearly a million type specimens, this is a huge task. It will be achieved, but not soon, unless the world is ready to support more taxonomists doing taxonomy. This in turn requires widespread acceptance of taxonomy as a mature and stimulating science.

S. Knapp, R. M. Bateman, N. R. Chalmers, C. J. Humphries, P. S. Rainbow, A. B. Smith, P. D. Taylor, R. I. Vane-Wright, M. Wilkinson
The Natural History Museum, Cromwell Road, London SW7 5BD, UK

1. Godfray, H. C. J. *Nature* **417**, 17–19 (2002).
2. Gewin, V. *Nature* **418**, 362–363 (2002).
3. Bisby, F. A. et al. *Nature* **418**, 367 (2002).
4. Moritz, T. *D-Lib Magazine* **8** (June 2002). www.dlib.org/dlib/june02/moritz/06moritz.html.
5. Lee, M. S. Y. *Nature* **417**, 787–788 (2002).

Metastasis: the role of chance in malignancy

Sir — The Concepts essay "A progression puzzle" by René Bernards and Robert A. Weinberg (*Nature* **418**, 823; 2002), demolishing a popular view of cancer metastasis, is welcome. To take their argument even further, the concept that tumours evolve by sequential selection of gene changes suggests that both malignancy and metastasis are mechanistically rather simple processes, and follow from a complete abolition of the signals that normally prevent cells from surviving and growing in the wrong place.

Bernards and Weinberg argue that the gene changes that make cancer metastasis possible cannot be fundamentally distinct from the preceding gene changes that occur as the cell evolves to form a primary tumour, and conclude that "genes and genetic changes specifically and exclusively involved in orchestrating the process of metastasis do not exist". If there is little difference between a primary malignant tumour and its metastases, the crucial issue becomes: what makes a tumour malignant (capable of metastasis)?

Applying the Bernards–Weinberg argument leads to the idea that malignancy

must also be a short step from earlier stages in tumour development. It can be separated by at most one unique gene event from overproliferation, and hence there is probably no more than one gene and genetic change per tumour that is specifically and exclusively involved in malignancy.

Malignancy can be defined either as the ability of a tumour to spread into the surrounding normal tissue and grow there to form a primary malignant tumour, or as the potential to form metastases. The two definitions are almost interchangeable because the one almost invariably implies the other. Pathologists observed this over a century ago and it is a cornerstone of clinical histopathology — and strong support for Bernards' and Weinberg's view.

Applying the Bernards–Weinberg selectability argument to malignancy itself, the transition from non-malignant to malignant can involve at most one gene change that is unique to the malignant transition, because cells cannot accumulate mutations that provide malignancy-specific properties (the ability to spread) during their evolution before they begin to spread. At most, one critical gene change could occur at the point at which malignancy begins.

This leads to a very simple view of malignant tumours and metastasis: malignant cells are cells that can grow in alien environments, and by a process that is not subtle, but largely random accident, they spread, first locally and then eventually through the body to form life-threatening secondary tumours. Metastasis is merely a rare, stochastic event: the chance escape of a cell into the vasculature, its arrival at a suitable site and its growth there.

Much has been made of the need for cells to cross basement membranes, but damage by trauma, inflammation or necrosis can breach basement membranes, and even normal somatic cells can cross vessel walls (S. Koop *et al. Proc. Natl Acad. Sci. USA* **93**, 11080–11084; 1996). Perhaps what is crucial is that when malignant cells are exposed to surrounding connective tissue they simply establish growth in the connective tissue space, instead of helping to repair the tissue organization by closing gaps or going into apoptosis (programmed cell suicide) when in the wrong place.

Paul A. W. Edwards

*Hutchison/MRC Research Institute,
University of Cambridge, Hills Road,
Cambridge CB2 2XZ, UK*

Metastasis: objections to the same-gene model

Sir — The model of cancer metastasis suggested by René Bernards and Robert A. Weinberg in their Concepts essay (*Nature* **418**, 823; 2002) is, in my view, a tautology. The suggestion that the same genes are exclusively responsible both for cancer-cell metastasis and for the emergence and proliferation of cancer cells is tantamount to saying “cancer cells that can proliferate do proliferate”. It would be a great loss if this type of idea caused a decline in research to investigate the existence of new genes involved in metastasis, a major factor in cancer mortality.

There is a finite probability that any cancer cell that can proliferate at one body site can also proliferate at other sites, if it can get there and stay there. How these two requirements are met is the real crux of the metastasis question. Bernards and Weinberg wish to dismiss the possibility that specific genetic changes, beyond those that govern proliferation, are required for successful cancer-cell relocation.

However, the argument Bernards and Weinberg used to arrive at this idea is, in my opinion, flawed. They argue that cells that acquire both proliferative and metastatic changes will be rare in primary tumours, thereby making it “difficult to

imagine how metastasis can ever proceed”. This is a remarkable proposition in the context of a discussion of the initiation of cancer cells, which is itself an extremely rare occurrence. The authors’ new concept — that metastasis is not due to a selected cell phenotype — would be better supported by a stochastic mechanism formulation.

Clinically, metastases range from presentation with large primary tumours to presentation without any identifiable primary tumour at all. It may be that the factors responsible for metastasis are cancer-cell proliferation plus myriad other small, unknowable variables that combine to create the conditions for relocation. This is a restatement of the tautology of Bernards’ and Weinberg’s argument.

James L. Sherley

*Biological Engineering Division, Massachusetts
Institute of Technology, Cambridge, Massachusetts
02139, USA*

Bernards and Weinberg reply — We appreciate the interest that our Concepts essay (*Nature* **418**, 823; 2002) has evoked. But Edwards misrepresents our thinking when he writes: “If there is little difference between a primary malignant tumour and its metastases, the crucial issue becomes: what makes a tumour malignant (capable of metastasis)?”. In fact, we argued that there are various types of primary tumour, some of which are preordained to become metastatic, others not. Hence, the differences lie between various distinct types of primary tumour. We suppose that in some tumours, the particular combination of alterations/mutations that enables cells to create a robustly growing primary tumour cell population also incidentally empowers them to become metastatic. Also, we do not imply that a single genetic change per tumour is involved in malignancy, as Edwards concludes.

We did not say, as Sherley asserts, that “the same genes are exclusively responsible both for cancer-cell metastasis and for the emergence and proliferation of cancer cells”. Instead, we argued that there are multiple alternative genetic pathways that lead to the creation of a primary tumour, each path being defined by the identities of the particular genes that are altered during tumorigenesis. According to our thinking, some combinations of genes that lead to primary tumour formation create growths that are unlikely to metastasize. Other combinations yield tumours that have a high proclivity for metastasizing. In the latter case, the combination of genes that yielded the primary tumour happens to be able to confer invasive/metastatic ability even though these phenotypes were not selected during the clonal expansions that

created the primary tumour mass.

We did not say, as Sherley asserts, that cells that acquire both proliferative and metastatic changes will be rare in primary tumours. Instead, we said that certain combinations of genetic alterations that are selected for the proliferative advantage they confer will, incidentally, also confer invasive/metastatic phenotypes. This is in no sense tautological. It is simply the statement of the possibly pleiotropic actions of certain cancer-associated genes.

René Bernards, Robert A. Weinberg

**Division of Molecular Carcinogenesis and
Center for Biomedical Genetics, The Netherlands
Cancer Institute, Plesmanlaan 121, Amsterdam,
The Netherlands*

*†Whitehead Institute for Biomedical Research
and Department of Biology, Massachusetts
Institute of Technology, Cambridge,
Massachusetts 02142, USA*

Chilean decree will save nights for star-gazers

Sir — Your editorial “Save starry nights” (*Nature* **418**, 709; 2002) states that Czechoslovakia is the only country with a national policy aimed at limiting light pollution. In fact, Chile, which now contains the largest concentration of optical telescope apparatus in the world, has taken a similar step.

The Association of Universities for Research in Astronomy, the European Southern Observatory and the Carnegie Institution worked for several years with the major Chilean universities and local and national government authorities to implement a strategy for controlling light pollution.

In December 1998, then-President Eduardo Frei signed a decree to establish an environmental norm regulating all outdoor lighting, not only in the areas surrounding the existing observatories, but in all of the II, III, and IV Regions of northern Chile, one-third of the total area of the country. The decree states in part: “The astronomical quality of the skies of the II, III and IV regions of our country constitute a valuable environmental and cultural patrimony recognized internationally as the best existing in the Southern Hemisphere for astronomical observations.”

The projected savings in energy costs from replacing polluting lights with well-shielded, energy-efficient ones should more than pay for the initial investment required for the changeover.

**Maxine Singer, Augustus Oemler,
Mark Phillips**

*Carnegie Institution of Washington,
1530 P Street NW, Washington, DC 20005, USA*

Move over Darwin

A look at the co-discoverer of natural selection. Neo-Wallaceism anyone?

In Darwin's Shadow: The Life and Science of Alfred Russel Wallace
by Michael Shermer

Oxford University Press: 2002. 368 pp.

\$35, £25

James Mallet

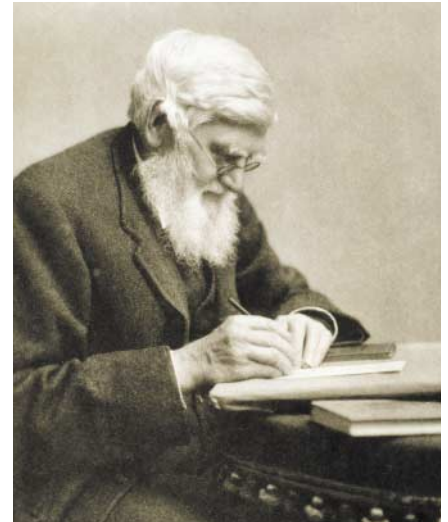
"In the beautiful forms of crystallization on his window [the scientist] recognizes the action of laws which may also have a part in the production of the similar forms of plants and many of the lower animal types. Thus the simplest facts of everyday life have for him an inner meaning, and he sees that they depend on the same general laws that are at work in the grandest phenomena of nature." Barely out of his teens in 1843, Alfred Russel Wallace here displayed a deep appreciation of the materialism and parsimony that led to his major discovery: natural selection. Fifteen years later, Wallace's letter to Charles Darwin from the Indonesian island of Ternate prompted a revolution in biology. After years of dithering and procrastination instead of publishing his theory of natural selection, Darwin found that Wallace had independently hit upon the same idea.

How could Wallace, the co-discoverer of the supreme materialistic theory of evolution, then be seduced by spiritualism, oppose vaccination, and argue that "an Overruling

Intelligence has watched over" human evolution and natural selection? Darwin was appalled: "You write like a metamorphosed (in retrograde direction) naturalist. And you, the author of the best paper that ever appeared in the *Anthropological Review*! Eheu! Eheu! Eheu! — Your miserable friend, C. Darwin." Wallace's paradoxical beliefs form the major subtext of this new biography. Michael Shermer, the editor of *Skeptical* magazine, clearly relishes the challenge of explaining Wallace's pseudoscience.

Wallace wrote that he "pondered much on the incomprehensible subjects of space, eternity, life and death. I think I have fairly heard and fairly weighed the evidence on both sides, and I remain an utter disbeliever." Religion, at least in 1861, could not explain Wallace's abandonment of the materialism he had announced 18 years earlier, and he never became conventionally religious.

Shermer follows earlier Wallace commentators, Charles H. Smith and the late Stephen Jay Gould, and interprets Wallace as a "hyper-selectionist". The capacities for art, music and philosophy are present even in "savage races", where they are not useful or beneficial, according to Wallace; therefore, they could not be explained by natural selection. Because these abilities resulted ultimately in the flowering of human culture,

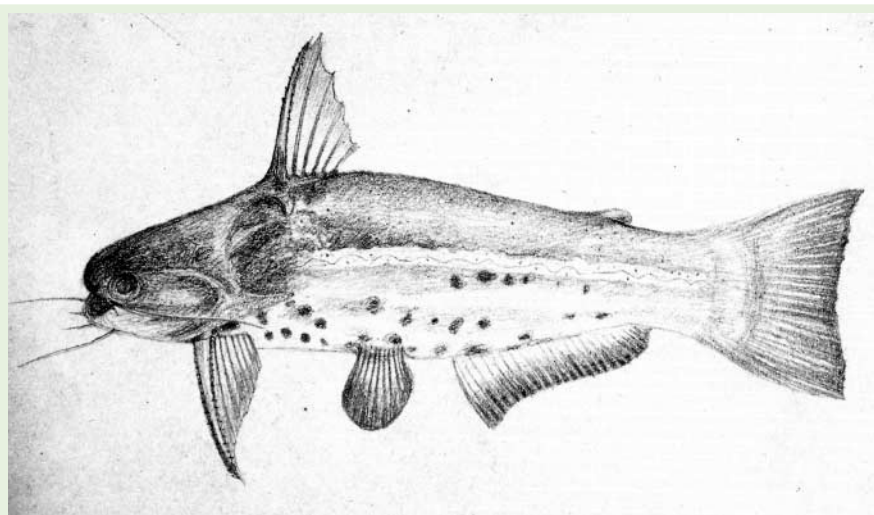


Change of heart: Alfred Russel Wallace abandoned natural selection for spiritualism.

Wallace felt that some sort of benevolent law, in addition to natural selection, must be directing human evolution. It was his strong belief in natural selection that made it supposedly logical that he should reject the principle for human origins. Also consistent was Wallace's membership of the anti-vaccination movement (vaccination would be necessary only if humans were not already optimal) and his support for human improvement by socialism and spiritualism.

The idea of Wallace as a hyper-selectionist is now decades old, dating from Gould's charming 1980 essay in *Natural History*. The obvious, almost forgotten but plausible alternative is that Wallace, like most of us, was simply inconsistent. His intellectual courage failed him, and he edged away from the yawning abyss into which the logic of materialism and natural selection led. He became a kind of theist, arguing that "intelligent beings, acting through natural and universal laws" directed human evolution for our benefit.

In contrast, dogged exploration of the limits of natural selection led Darwin to explain such oddities as rudimentary organs and the apparently non-adaptive and flamboyant colours of male animals. Today, Darwin's viewpoints on these topics are completely vindicated: for example, the idea that peacock tails evolved by female choice is strongly supported by experiment. After the 1860s, Wallace could never stomach such a quirky kind of evolution, and argued strongly against it. But Darwin wasn't always right: he erroneously accepted the inheritance of acquired characters (Lamarckism) in evolution, whereas Wallace rejected it. The problem here was the data. Wallace's zeal for natural selection



Survival of the fishes

Alfred Russel Wallace's first large expedition to South America nearly ended in disaster. After four years collecting specimens up the Rio Negro and Rio Uaupés, Wallace was returning home when the ship he was travelling on caught fire. He escaped with only a tin box containing a few personal belongings and a collection of drawings

of fishes and palms — the rest of his extensive collection was lost forever. Now Mônica de Toledo-Piza Ragazzo has organized the first publication of this collection of drawings, one of which is reproduced here, in *Fishes of the Rio Negro: Alfred Russel Wallace* (University of São Paulo, 120 Brazilian reals, US\$50).

probably explained his aversion to lamarckism, despite the evidence, which apparently supported it; Darwin believed the faulty data.

Shermer covers all the other aspects of Wallace that one wishes and expects. A good sceptic, Shermer carefully debunks conspiracy theories found in other biographies, such as that Darwin and his friends unfairly took the credit for natural selection from Wallace. Shermer uses new manuscript information to show that such a plot is highly unlikely, and that Wallace strongly approved of the "gentlemanly arrangement" to co-publish with Darwin in 1858 as soon as he heard of it. Indeed, Shermer makes it clear that the young Wallace's reputation and career benefited greatly from the joint publication of natural selection in the *Proceedings of the Linnean Society*.

Shermer is weaker on the background, particularly on species and the history of evolutionary ideas. Much of it seems derivative of Ernst Mayr's *Growth of Biological Thought* (Harvard University Press, 1982). Shermer unquestioningly accepts Frank Sulloway's statistical analyses, putatively showing that "later-borns" such as Wallace are more radical scientists than first-borns. This seems oddly uncritical for the editor of *Skeptic*, and it is easy to think of exceptions.

Shermer's description of Wallace's life is excellent, and the bibliography of his publications and manuscript sources is the most complete to date. Shermer befriended the Wallace heirs and accessed their archive to gain new insights into their ancestor (this collection has since been acquired by the Natural History Museum in London). The book is full of interesting anecdotes, for instance that Wallace made a bet with a flat-Earther, John Hampden, and won £500 by means of a clever experiment on the Old Bedford Canal. Sadly, Wallace had to spend more than that on lawyers to recover the money from his opponent, who was clearly not a good loser.

Biographers and historians like to find a new angle on their subject, a selling point for their work that lasts until the next treatment comes along to disprove it. But wishful thinking has led to a lot of nonsense being published about Wallace, and about his supposed conflict with Darwin in particular. When it comes to Wallace's travels, discoveries, public life and beliefs, truth is at least as strange as fiction, and it is a relief that Shermer entertains the reader by getting the facts of Wallace's life right. Let's hope that most of the contrarian ideas that can be twisted out of Wallace's life have already been aired. But as the spate of interest in the history of evolution continues, more new Wallace biographies are on the horizon. I wonder what the next angle will be? ■

James Mallet is in the Department of Biology, University College London, Wolfson House, 4 Stephenson Way, London NW1 2HE, UK.

Mapping biodiversity

World Atlas of Biodiversity: Earth's Living Resources in the 21st Century

by Brian Groombridge & Martin D. Jenkins
University of California Press: 2002. 340 pp.
\$54.95, £37.95

Kevin J. Gaston

The sheer pervasiveness of human activities is hard to envisage. Conventional maps of the distribution of the present 6 billion or so humans illustrate the peaks and lows of those presently residing on the major landmasses. But they fail to capture the large numbers of people who may pass through areas that hold no or rather small resident populations, the many thousands of individuals who at any given time are dispersed across or under the oceans (in ships or oilrigs, for example) or in the atmosphere (in aircraft), or the handful that are out in space. Track all of these activities over even a moderate period and much of the Earth would have been criss-crossed innumerable times.

Humans are the most widely distributed 'terrestrial' vertebrate, vastly more abundant than other species of such body size. We operate an enterprise of bewildering magnitude and complexity to sustain these individuals and the societies to which they belong. It is hardly surprising that in the face of these pressures, much of the variety of life on this planet, be it species, populations or genes, is being steadily expunged. What is more baffling is that apparently so few people have grasped the magnitude of the

problem and the breadth of its implications. On closer reflection, the media attention that is paid to these issues, and the general availability of appropriate resources (articles, books and television programmes, for instance), is considerably less than for many other fields of significant human concern. Those of us attempting to maintain biodiversity should not be beguiled into generalizing from the state of our own awareness, nor into overestimating media coverage just because we pay attention to it when it does occur.

This book by Brian Groombridge and Martin Jenkins of the United Nations Environment Programme's World Conservation Monitoring Centre sets out to provide an overview of the current state of global biodiversity. The authors try to step into the breach and make the facts more accessible to a wider readership beyond the community of professional biologists. The book is effectively a second edition of a volume published two years ago, but it has been much extended and updated, with a hopeful eye to renewed interest in the area being spurred by the recent Johannesburg World Summit on Sustainable Development.

The volume is well written, and there is little with which to disagree. The coverage is reasonably comprehensive and remarkably up to date, with many of the sources cited having been published in the past 18 months. The first two-thirds of this book comprises chapters addressing, in some detail, the nature and origins of biodiversity, its dynamics and distribution, and its relationship to humans both now and in the future. These pages are stuffed full of striking photographs, tables, graphs and maps. The value of some of these is perhaps a little



The mist rises: protecting biodiversity means raising awareness about the issue among policy-makers.

questionable, and others are unfortunately divorced from the caveats and cautions made explicit by their originators. But most are succinct and informative, with a number of the maps being original and worthy of being brought to the attention of researchers in this field.

The final third of this book comprises six appendices detailing, respectively, the phyla of living organisms, important food crops, domestic livestock, recent vertebrate extinctions, biodiversity at country level, and important areas for freshwater biodiversity. Much of this material is rapidly becoming staple for volumes on biodiversity, and is available elsewhere, so one has to question the value of its repeated inclusion (accepting that some of it receives frequent revision). This, of course, rests on who the readership is likely to be. There are many people around the world who would learn much that is of importance from reading this book, including educators, politicians, policy-makers and the general public. But it must do more than simply get onto their shelves, it needs to be well thumbed. If the authors and publishers can achieve this, they will have done all 6 billion of us a great service. ■

Kevin J. Gaston is in the Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK.

Mathematics off the main line

Mathematics Elsewhere: An Exploration of Ideas Across Cultures

by Marcia Ascher

Princeton University Press: 2002. 224 pp.

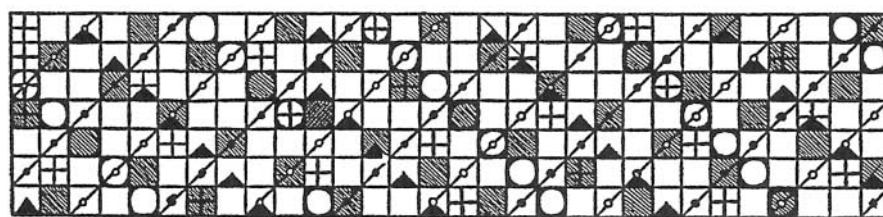
\$24.95, £17.95

John O'Connor

Mathematicians like to trace the ancestry of their subject back to the Renaissance and then, with a bit of disruption due to the Dark Ages (and some help from Islamic and Indian mathematicians), back to the Ancient Greeks, and perhaps even earlier to the Babylonians, Egyptians or the oldest Middle Eastern civilizations that we know of. More recently, however, we have become aware that many of the important ideas that we have placed on this main line of development were discovered independently in a variety of other societies.

Marcia Ascher's latest book provides yet more evidence that mathematical ideas that Western mathematicians have regarded as recent sophisticated constructs also have roots in geographically diverse cultures. She considers several different areas where such independent development has taken place.

Calendars offer an insight into how vari-



Marking time: the Balinese calendar, here represented on paper (top) and wooden tika, requires complex calculations.

ous societies have attempted to resolve the problems associated with the different length cycles of Sun, Moon and seasons. We feel at home with our Western calendar in its present gregorian variant, and it is only when we contemplate something like the rule for locating Easter that we realize just how much calculation it requires. The mathematician Carl Friedrich Gauss developed the system that we now use for modular arithmetic to handle the problems associated with calculating Easter. More recently, the version of the formula printed in *The Book of Common Prayer* was one of the influences that turned Bertrand Russell towards mathematics.

In *Mathematics Elsewhere*, Ascher considers a variety of other calendars: two from South Pacific cultures, as well as the Jewish, Mayan and Balinese versions. All of these lead to extremely intricate calculation problems. The Balinese calendar, for example, has ten weeks of different length running simultaneously, each with its own labelling system for days. The problem of moving backwards and forwards between these different systems would tax the mental-arithmetic capacity even of those of us raised to calculate in pounds, shillings and pence, let alone those from the decimal-currency and calculator age. As the author points out: "Some of the mathematical ideas and questions are restricted to specialists, but, nevertheless, the logic of interlocked cycles is pervasive in the cultures."

Other 'modern' ideas crop up in unlikely places in tribal cultures. The divination rituals practised in the Caroline archipelago in the Pacific, in some areas of Nigeria and in the island of Madagascar have common elements related to the mathematics of boolean algebra, and involve algorithms only recognized as such by twentieth-century mathematicians. Most of these processes operate by

a process of binary choices and, as Ascher explains: "It is, therefore, not surprising that many, if not most, of the various modes of divination have mathematical ideas as an integral component."

One of the central ideas of modern mathematics is the idea of a 'relation'. Two objects can be considered 'equivalent' if it is convenient to regard them as the same for some purposes. So topologists like to think of their spaces as the same if they are related to each other in the appropriate way. This process of blurring the distinction between related mathematical objects is a very modern concept, but the same idea has occurred independently in many different cultures. So, for example, in one of the most interesting investigations in this book, the maps or diagrams made by the sailors of the Marshall Islands in the Pacific to represent the way that ocean currents circulate around their archipelago do not represent the islands in the usual geographical way, but according to how their experience in navigating with water and waves enables them to locate their position.

In a completely different context, certain Basques, some of the Tongans of Polynesia and a subset of the inhabitants of Ethiopia use the same basic mathematical formulation to represent the social ranks, divisions and connections in their societies.

Mathematics Elsewhere is a useful reminder of how universal mathematical and logical structures are in any culture. Mathematicians will enjoy seeing the subject they love cropping up in apparently unexpected contexts. Non-mathematicians should be encouraged to realize that some of the processes that seem to appear naturally in everyday life do in fact have a mathematical content — so, just as Molière's Monsieur Jourdain was talking prose without realizing it, they are in fact doing mathematics without being aware of it. ■

John O'Connor is at the Mathematical Institute, University of St Andrews, St Andrews, Fife KY16 9SS, UK.

The ocean's veil

Victor Smetacek

Imagine being whisked in a flight of fantasy into the microbial world of the oceans' sunlit layer, and being able to see the smallest organisms on the planet running its largest ecosystem. At your new, micrometre-sized scale, water is so thick and viscous that attempting to swim through it would be futile. Yet some peculiar-looking bacteria and other strange organisms would be zooming through the honey-thick medium with effortless ease. It is too early to write a guidebook on life in the microbial environment, and such a text would soon be outdated, because surprises continue to pour in.

Just 25 years ago, the physics of this environment were revealed for the first time by E. M. Purcell in a delightful lecture to a gathering of physicists (he called bacteria "animals"). The habitat he sketched, with his collage of equations, is a wonderland so radically different to our sensory experience that intuition seems only to be right for the wrong reasons.

Viscous forces rule micrometre-sized organisms just as the force of gravity rules those at the large end of the size spectrum. We, as large animals, tend to equate 'viscous' with 'slow', but in actuality the microbial realm is full of activity. The medium is continuously and thoroughly mixed by molecular diffusion, supplying dissolved nutrients to cell surfaces with such relentless efficiency that it matters little whether a cell stays put or moves around at a few body lengths per second. To

experience a change of surroundings, a bacterial cell has to race against its medium with the speed of a greyhound — about 20 body lengths per second. This is the speed attained by our gut bacteria (*Escherichia coli*), yet some planktonic bacteria can move five times faster.

These motile bacteria are propelled by rigidly rotating flagella, which are driven by reversible rotary motors. Purcell calculated that rotating this propeller costs surprisingly little energy and, in a viscous world, bumping into an object causes no damage. So bacteria can afford to zip around at their maximum speed, restlessly searching for nutritive hotspots such as leaking large organisms or flocks of detritus. But many bacteria are non-motile and thrive on nutrients supplied by molecular diffusion.

Planktonic microbial communities — comprising bacteria, viruses and the smallest protists (eukaryotes less than 5 μm in size) — which inhabit the surface layers of all water bodies, run their food webs in a strikingly similar manner. The bulk of the 'plants' (cyanobacteria and algae) and 'animals' (protozoa) are represented by relatively few genera from several ancient, widely separated lineages, suggesting that extant species within the food web have coevolved. Although the species composition varies, the structure of the food web in terms of the relative proportions and the maximum abundances of the 'plants', 'animals' and bacteria is remarkably conserved, contrasting strikingly with the huge variation in biomass and abundance of planktonic organisms larger than about 10 μm .

When resources are plentiful, microbial growth rates increase significantly, but cell numbers do not exceed a threshold value. The remarkably stable bacterial numbers — ranging around a million cells per millilitre — are explained by the heavy predation pressure of a few genera of flagellated protozoa that hunt, capture and ingest bacterial cells individually. If a bacterial population manages to grow through this predator gauntlet, it is decimated by viruses. So the numbers of bacteria, and apparently also of other microbes, are kept within boundaries by the double lock of predators and pathogens. As in an overgrazed pasture, potential resources are underexploited.

Where predation pressure is heavy, avoidance and defence mechanisms will be selected for — planktonic bacteria hide by forming spores, grow large by forming filaments, produce tough cell walls, secrete toxins and can also flee at high speed. Successful escape reactions can be triggered at encounter, during handling and even after ingestion. So not only do motile bacteria race to find hotspots, but they are also difficult

Microbial food webs

These tightly regulated oceanic communities consume small particles but let larger ones sink into the depths below.

to catch. As their wake is obliterated almost instantaneously by diffusion, they cannot be tracked by predators. But much larger particles (the nutritive hotspots) can be tracked, so bacterial populations growing on or inside them are vulnerable to attack. Many bacteria will be eaten; others will escape from the hotspot. Thus, the rate at which bacteria break down large particles will be slowed by the presence of predators.

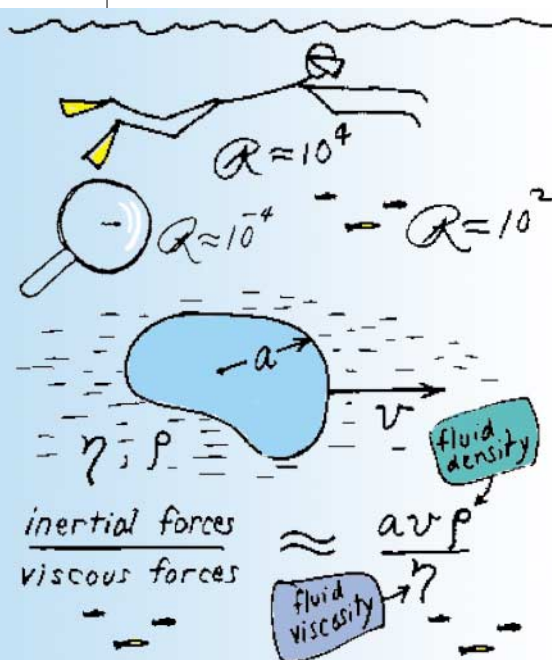
Apparently, the stability of bacterial numbers is maintained by a set of interacting biophysical 'rules' that balance the growth and loss rates of cells within the microbial size range. As one ascends the food chain, these rules weaken with increasing organism size — the ciliate predators of microbial organisms can increase their populations by orders of magnitude. So microbial production is channelled up a food chain that is tethered at its base. The excess nutrients are taken up by large algae that end up in large zooplankton. In all cases, the end result is large particles that sink from the surface layer faster than they can be broken down by bacteria. This rain of large particles impoverishes the productive layer and provides food for the deeper-living organisms below, with immense consequences for aquatic ecology, ocean biogeochemistry and global climate.

The microbial network that stretches like a veil across the watery face of the planet functions as a self-restrained, semi-permeable membrane that retains small particles but lets bigger ones pass through. A closer look at the rules of the game in the fast lane of the microbial realm would require an *in situ* computerized telemicroscope. Could such an instrument do for microbial ecology what Galileo's telescope did for astronomy? After all, there are orders of magnitude more bacteria in the ocean than there are stars in the Universe. ■

Victor Smetacek is at the Alfred Wegener Institute for Polar and Marine Research, Am Handelshafen 12, 27570 Bremerhaven, Germany.

FURTHER READING

Copley, J. *Nature* **415**, 572–574 (2002).
Purcell, E. M. *Am. J. Phys.* **45**, 3–11 (1977).
Fenchel, T. *Science* **296**, 1068–1071 (2002).
Finlay, B. J. *Science* **296**, 1061–1063 (2002).
Matz, C. *Mechanisms of Grazing Resistance in Aquatic Bacteria*. Thesis, Univ. Kiel (2002).



Wonderland: equations sketched by Purcell in his lecture on the physics of the microbial world.

Life with the artificial Anasazi

Jared M. Diamond

How can the complex dynamics of human societies — such as population rise and fall, and movement — be explained? Combining masses of data with computer modelling is a fresh way forward.

Many complex systems exhibit properties that cannot be predicted by verbal reasoning, or even pure mathematics, from the behaviour of their components. Human society is no exception. Take, for example, the fact that “a traffic jam, which results from the behavior of and interactions between individual vehicle drivers, may be moving in the direction opposite that of the cars that cause it” (page 7280 of ref. 1). A promising approach to this general problem is ‘agent-based modelling’, which populates specified spatial landscapes with agents that assess the landscape and behave according to a set of decision rules^{1,2}. Computer calculations reconstruct, step by step, the trajectory of the model, and compare it with real history to test the accuracy of the input information and the model’s assumptions.

Two papers^{3,4} have set new standards in archaeological research by applying this approach to a historical problem: the rise and fall of the Anasazi civilization of Native Americans in the southwestern United States. The authors have combined quantitative information on environmental fluctuations with rules of behaviour expected for Anasazi households, to calculate an Anasazi historical ‘trajectory’ in unprecedented detail. As with any good modelling program, two erroneous predictions of a first-generation model helped to identify missing factors that were then incorporated into improved models.

Until its disappearance around AD 1350, Anasazi society was marked by fluctuations in total population size, mean settlement size



Figure 1 Home for the Kayenta Anasazi. This view of Long House Valley, northeastern Arizona, shows the fertile valley floor flanked by unfarmable upland zones.

and preferred habitat. But why? It is generally agreed that the answers depend partly on rainfall and on groundwater levels, variations in either of which would affect Anasazi maize-based agriculture. Proxy measures of both of these variables are available for a group called the Kayenta Anasazi of Long House Valley in northeastern Arizona (Fig. 1), together with archaeological surveys of

the numbers and distribution of houses for the entire span of the valley’s Anasazi occupation. But how can one assess whether the proxies can explain the survey results? To give just one hint of the complexities involved: readily available groundwater can make rainfall unnecessary for maize agriculture, but the dependence of maize growth on rainfall and groundwater levels varies among

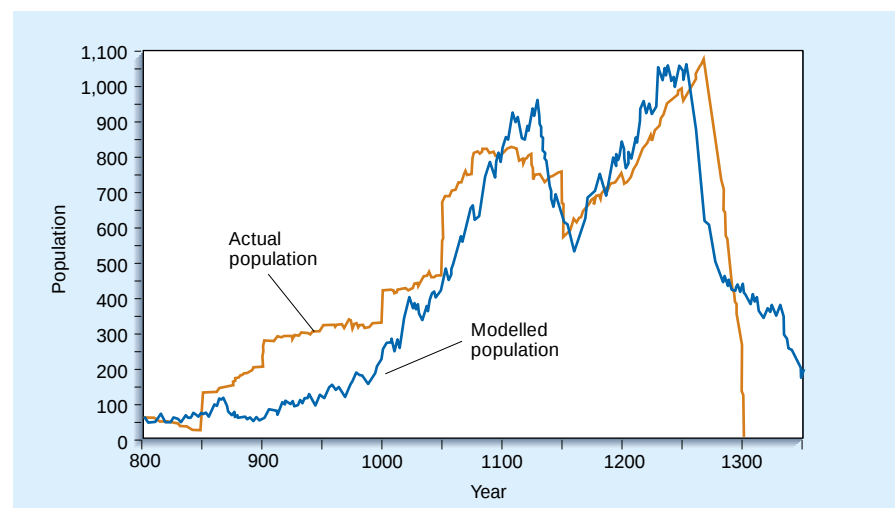


Figure 2 Ups and downs of the Anasazi in Long House Valley. Shown here are actual (orange) and modelled (blue) numbers of the valley’s population from AD 800 to AD 1350. Actual population is based on the number of house sites at a given time, assuming five people per house. The fluctuations in the modelled population reproduce those of the actual population fairly well. The main discrepancy is that the Anasazi completely abandoned the valley by 1300, at a time when the model predicts that it could still have supported a substantial population. (From Fig. 2 of ref. 4.)



100 YEARS AGO

At the festivities held in Bologna on the occasion of Mr. Marconi's return to his native town, Prof. Augusto Righi, in congratulating his former pupil in his successes, spoke to the following effect:— Perhaps no one can appreciate better than I his exceptional inventive power and his unusual intellectual gifts. I remember with great pleasure his visits when quite a young man, for asking my advice, for explaining his experiments, made with simple apparatus ingeniously put together, and for keeping me informed of his new projects, in which his passion for applied science always stood out. Even then I predicted that he would sooner or later attain fame. The system of wireless telegraphy which he derived from Hertz's classical experiments ... is the most pleasing transference to the field of practical industry of those instruments and principles which might have seemed to be relegated to the domain of natural philosophy. ... It is to the credit of Marconi that he has once more proved how much those are in error who regard with disdainful or indifferent eyes the work carried on continuously in the silence of the laboratory by the modest and disinterested scientific students, and who only appreciate science in proportion to the immediate uses that can be obtained from it. From *Nature* 9 October 1902.

50 YEARS AGO

Philosophic Problems of Nuclear Science. One summer evening in 1925, a small and very select meeting of theoretical physicists took place in one of the Fellows' rooms in Trinity College, Cambridge. It was a special occasion. ... A brilliant young mathematical physicist, fresh from Munich, Göttingen and Copenhagen, was about to expound his work in the field of atomic physics to a handful of people capable of appreciating it. The speaker was ... Werner Heisenberg, before long to become a celebrity for his 'uncertainty principle'. Subsequent events have added not only to his distinction but also to the debt which philosophers and scholars everywhere owe him for the profundity of his thought, and for the elegance of his expression. The present book does something to bring these things home to English-speaking readers. 'Something' is probably fair comment, for here is a collection of lectures, originally in German, translated with obvious sincerity but not always with complete idiomatic success. From *Nature* 11 October 1952.

the valley's six habitats used for agriculture.

The first-generation model (by Dean *et al.*³) had three steps. First, the width of annual tree rings depends on that year's climate, especially rainfall and temperature. Tree rings provide a climate record for Long House Valley extending from the present back to AD 200, with no gaps. To convert tree rings into maize yields, the authors used an empirical relation, observed over the last century, between tree-ring width and a climate measure (the Palmer Drought Severity Index, PDSI)⁵ much used by agricultural scientists to calculate crop yields. The modern relation between maize production and PDSI is known for 55 different soils in southwestern Colorado, and one of the 55 closely matches the soil of one of the six Long House Valley agricultural habitats (Long House Valley General Valley floor soil). By converting tree-ring width to PDSI, and then PDSI to maize yield, Dean *et al.* calculated rainfall-dependent maize production, in kilograms per hectare, for that one soil for each year from AD 800 to AD 1350.

The next step involved the reconstruction of past rises and falls of groundwater from flood-plain stratigraphy: rising groundwater is associated with sediment deposition when the flood plain is well vegetated and stable; falling groundwater is associated with instability and erosion, which produces gulleys called arroyos. Soil stratigraphic units, and thus groundwater trends, can be dated to within a few decades by the styles of Anasazi pottery, radiocarbon-dated materials and germination dates of trees buried in each unit. It is known roughly how the combined effect of rainfall and groundwater on maize yields differs among the valley's six habitats. So Dean *et al.* took the time sequence of mean maize yields calculated in step 1 for General Valley floor soil, modified it to estimate the time sequence of yields for the other five habitats, and incorporated random spatial and temporal variations in yields around those mean values.

The last of the three steps in the first-generation model was to construct an 'artificial Anasazi' history by computer. This involved dropping a few virtual Anasazi farmers into the valley at AD 800; feeding them each year with the calculated maize crop; and letting them bear and feed children, grow old, move house sites, and send off grown children to build new houses, according to rules observed for recent maize-growing societies of Pueblo Indians descended from the Anasazi. Examples of these rules are that annual maize consumption per person is 160 kg, that surplus maize can be stored for up to two years, and that children marry and move out at age 16. The iteration was carried out for each year in turn, depending on the calculated maize crop.

The result is one estimation of the valley's total population, and the spatial distribution

of its house sites. Dean *et al.* did many such simulations for each year. They turned out to differ in detail (because of the built-in stochastic variation in local maize yields) but were broadly similar. The simulations were then averaged and compared with the actual Anasazi population and its spatial distribution, as deduced from an archaeological survey of house sites (dated by their pottery styles) over the whole of the valley.

The first-generation 'artificial Anasazi' society and the actual Anasazi society compared well in several respects: population rise and fall, population shifts among the six habitats, and alternations between a few large settlements and many scattered houses. But there were two exceptions.

First, the simulated population was about six times the actual population. A likely explanation for this failure lies in previous experience with agent-based modelling, which suggested the model's deterministic demography as the flaw — for instance, that every woman became a mother at precisely age 16, and every household lasted for 30 years. When stochastic variation around those mean values was introduced, the resulting second-generation model⁴ accurately reproduced the population size as well as its trajectory with time (Fig. 2). For example, the population peak around AD 1250 was 1,070 people actual, 1,040 simulated. Overall, putting Fig. 2 into words, the population first rose in an initially under-used landscape, then remained flat between about AD 900 and 1000 because of groundwater limitation; it rose from 1000 to 1130 and again from 1180 to 1270 because of high groundwater and high or at least constant rainfall; and it crashed from 1130 to 1180, and again after 1275 because of a coincidence of drought with falling groundwater.

The other model failure comes after AD 1275, when the coincidental water shortages induced the actual Anasazi to abandon Long House Valley completely. Where did they go? Many people may have died of starvation. But others moved south to the Hopi pueblos and other climatically less-stressed areas, as shown by population surges and the appearance of Long-House-Valley-like ceramics and sites there. But the model suggests that valley maize yields after AD 1275 could still have fed 400 people, almost half of the peak population earlier in that century. For the first two centuries of their occupation of the valley, the Anasazi were content to live there at a population of under 400. Why did everybody leave the valley after AD 1275, when it seems that half of them could have continued to feed themselves?

When the Anasazi abandoned their settlements, the survivors were evidently motivated by something not considered in the model. One possibility⁴ is that complex human societies require a certain population size to maintain institutions that its

citizens consider to be essential. How many people would choose to remain in London if half of their family and friends had just starved to death there, and if the tube, pubs and churches were no longer open? Another possibility is that although the environments of both AD 1275 and 1050 could feed 400 people according to the model, they differed in ways not taken into account by the model. For instance, perhaps the forests still growing near the valley in 1050 had been felled by 1275, leaving no nearby timber for buildings and fire, or perhaps soil fertility became exhausted.

One overall message of this project lies in the saying, "God is in the details". Without all those data on rainfall, groundwater, soil types, crop yields and household behaviour, verbal claims such as "The Anasazi abandoned their homeland because of drought" are as likely to be wrong as right. Another message (even to a computer-phobe like me) is that the trajectories of human societies are so complex that computer modelling is essential for evaluating the consequences of the data. Finally, models involving many input parameters provoke much scepticism. But here the input parameters are measured

ones, and the model is really nothing more than a calculation of the consequences of those input parameters, given certain assumptions.

Experience with the 'artificial Anasazi' shows that confronting a computer model with reality yields conclusions about the questions that interest any archaeologist: why populations increased, why people moved, why their settlements varied in size, and why their society finally disappeared. This study should inspire archaeologists to gather the masses of data required to test their own verbal claims about human societies in other times and places.

Jared M. Diamond is in the Department of Physiology, University of California Medical School, Los Angeles, California 90095-1751, USA. e-mail: jdiamond@mednet.ucla.edu

1. Bonabeau, E. *Proc. Natl Acad. Sci. USA* 99, 7280–7287 (2002).
2. Kohler, T. A. & Gummerman, G. J. *Dynamics in Human and Primate Societies: Agent-based Modeling of Social and Spatial Processes* (Oxford Univ. Press, New York, 2000).
3. Dean, J. S. et al. in *Dynamics in Human and Primate Societies: Agent-based Modeling of Social and Spatial Processes* (eds Kohler, T. A. & Gummerman, G. J.) 179–205 (Oxford Univ. Press, New York, 2000).
4. Axtell, R. L. et al. *Proc. Natl Acad. Sci. USA* 99, 7275–7279 (2002).
5. Palmer, W. C. *Meteorologic Drought Res. Pap.* 45 (US Weather Bureau, Washington DC, 1965).

Superconductivity

Putting the squeeze on lithium

N. W. Ashcroft

At low temperature, some elements are superconductors under normal pressure. Others become superconducting if the pressure is raised. Lithium is the latest low-temperature, high-pressure superconductor.

The report by Shimizu *et al.*¹ (page 597 of this issue) of superconductivity in lithium, a light metal in group I of the periodic table, is noteworthy for being at odds with an old historical imperative in superconductivity. This remarkable phenomenon was discovered in 1911, and over the following 40 years a considerable database of superconductors was amassed that included both elements and compounds. Its scope was sufficient to invite a search for signs of systematics, and in 1955 Bernd Matthias² summarized the situation with a plot of the superconductors' transition — or superconducting — temperature as a function of the average number of valence electrons. One of its more dispiriting features was the apparently resolute exclusion of the putatively 'simple' metals of valence 1.

The half-century since then has seen a veritable avalanche of new superconductors identified, including the extraordinary but still somewhat mystifying class of high-temperature cuprate superconductors, and it became clear that attempts to find systematic effects would require far more searching assessments than that of Matthias. Never-

theless, suspicions that valence-1 elements would not be part of the unfolding story of superconductivity seem to have lingered on.

This past half-century has produced another significant development, namely the relentless drive to subject elements and compounds to ever-increasing static compression, in some cases approaching an order of magnitude above normal densities. Among other achievements, these pressures have succeeded in turning non-superconductors (and even non-conductors) into superconductors. The success of Shimizu *et al.*¹ now brings the tally of pressure-induced superconducting elements to 23 (Fig. 1). This is a very considerable achievement considering that at ordinary pressures the number is just 29.

The changes wrought by pressure are often striking. To make the point, if conditions of one million atmospheres or so were normal on Earth, some characteristic features of the periodic table would look very different from the present one-atmosphere construct. In particular, sulphur, in group VI, would be a metal, and it would also be an exceptional elemental superconductor, with a transition temperature reaching 17 K (ref. 3). Once again this is a case of historic note, for valence-6 elements were also excluded in Matthias's early one-atmosphere systematization.

So now we have the important diamond-cell experiment on compressed lithium, a group-I element, which Shimizu and colleagues¹ find to be superconducting. At pressures as high as 48 gigapascals (roughly 480,000 atmospheres), the transition temperature rises to 20 K. Notable though it is, the discovery is not entirely unheralded, because in 1986, with a quite different apparatus, Lin and Dunn⁴ reported resistance changes in lithium at high pressures and low temperatures, and in interpreting their results they did not rule out the onset of a superconducting state.

Lin and Dunn apparently had no easy way

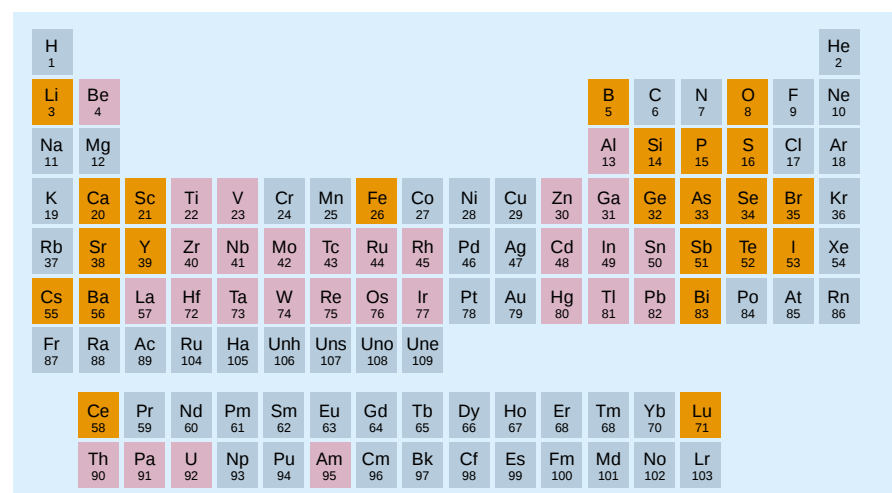


Figure 1 Superconductors under pressure. The colour code of this periodic table (adapted from ref. 13) shows elements that superconduct under normal, atmospheric pressure conditions (purple) and those that superconduct when subjected to high pressure (orange). Shimizu *et al.*¹ confirm the superconductivity of lithium at high pressure, bringing the number of such elements to 23. Under normal pressure conditions, 29 elements are superconductors.

of applying a magnetic field in their experiment. But, for many workers in this area, a *sine qua non* for establishing a new superconductor is the observation of the Meissner–Ochsenfeld (MO) effect — when an element or compound enters the superconducting state, magnetic flux that previously permeated the sample is suddenly excluded. Also, in certain superconductors, applying a magnetic field of sufficient magnitude can suppress superconductivity, and this is exactly the effect seen by Shimizu and colleagues. Although they also hold observation of the MO effect to be “indispensable” and were unable to achieve this, their demonstration of this suppression is nevertheless an important pointer.

Another indication of the onset of superconductivity is a drop in electrical resistance, which Shimizu *et al.*¹ observed. This, however, is a slightly indeterminate procedure, as can be seen from the authors’ data, and future experiments may well lead to some minor revisions of the transition temperature they have measured. Nevertheless, the evidence seems compelling that a superconducting state in lithium has been attained. As

the authors note, the reported transition temperature makes lithium (for now) the element with the highest superconducting transition temperature — with, so to speak, not a *d*-electron in sight.

Although it may be regarded as a simple element in atomic terms, in structural terms lithium in its condensed state at high pressures is complex. It seems to be the case⁵ (and indeed it was predicted⁶) that, under pressure, lithium takes on structures quite unlike the familiar body-centred-cubic phase, including structures that have an even number of atoms in the unit cell. How this structural complexity connects with the theory of superconductivity is an intriguing question, for perhaps one of the most difficult theoretical problems in the field of condensed-matter physics is accurately predicting superconducting transition temperatures. Nevertheless, general arguments (based on the density of states, the strength of the electron–phonon interaction, the scale of the Coulomb pseudopotential, and so on) can bolster the expectation that a state of significant superconductivity will occur, and did in fact make a strong case for lithium⁶.

But calculations⁷ of lithium’s transition temperature seem to come up with a value that is a factor of four or so above Shimizu and colleagues’ measurement. In a way, this discrepancy could be seen as a propitious opportunity — the pressure experiments on lithium and on other light-element systems may at last offer the prospect for systematic studies of the behaviour of the dynamic electron–electron interaction and in particular of the Coulomb pseudopotential, the manifestation of direct electron–electron repulsion that generally works against electron pairing (unlike the electron–phonon interaction, which in lithium probably promotes it).

It may be that the early Matthias rule should simply be rephrased to read that single-band, rather than single-valence, elements are in general not favoured for superconductivity, for there is accumulating evidence that structures resulting in multiple bands are predominant among superconductors. If structural anisotropy in layered structures is a route to higher-temperature superconductors, as Ginzburg and Kirzhnits have suggested⁸ and as seems to be the case with the superconducting latecomer MgB₂

Behavioural ecology

Excuses for avian infidelity

Over the past decade, molecular methods have provided ample evidence that, in most apparently monogamous animal species, it is quite common for a couple to raise offspring that resulted from either the male or the female mating with another partner. This is particularly evident in birds. Yet explanations for why some female birds actively seek copulations with other males, or why males mate with other females and allow them to lay eggs in the male’s nest, are speculative. Elsewhere in this issue, however, a group of ten biologists from six countries provide convincing evidence that, in some cases, the phenomenon can be explained by how closely related the male and female of a pair are (D. Blomqvist *et al.* *Nature* **419**, 613–615; 2002).

The authors found that in three shorebird species, including the western sandpiper (*Calidris mauri*), pictured here, the birds of a pair that is rearing ‘illegitimate’ offspring are more closely related to each other than the average relatedness. So it seems plausible that they sought alternative mates to reduce the harmful effects of inbreeding. Relatives often carry the same versions (alleles) of various genes,

so there is a strong chance that their offspring will inherit two identical alleles — one from each parent — of those genes. This matters for recessive alleles, which are expressed only if both copies are the same.

Increased expression of recessive alleles depends on the degree of relatedness of the parents and the frequency of the alleles in the population. The effect is greater for rarer alleles — and one would expect alleles with harmful effects to be both rare and recessive. This means that inbreeding is usually associated with increased expression of harmful alleles and a low survival rate of offspring, and therefore that mating with relatives should be avoided.

So there is a logic to Blomqvist and colleagues’ finding that shorebirds that are, through some social constraint, paired to a relative then seek alternative partners. The strategy would increase the fitness of offspring and could be evolutionarily selected for. The big question now is, how do the birds know who their mate is? It is likely that many birds can recognize individuals that grew up in the same nest, but we have no idea how they



recognize half-siblings from other nests, or cousins. In rodents, smell provides clues, but in most birds the sense of smell is comparatively underdeveloped.

However, Blomqvist *et al.*’s observation fits with other evidence that birds are cleverer and learn more from past experiences than previously thought. For instance, birds can assess the breeding success of other individuals from the same species, and use that information in deciding where to breed the following year (see, for example, B. Doligez *et al.* *Science* **297**, 1168–1170; 2002). They can also adjust the time that they reproduce according to past

experience (F. Grieco *et al.* *Science* **296**, 136–138; 2002). All of these findings provide better insight into bird ecology, but also sound a cautionary note to researchers: the idea that birds react to, and compensate for, experimental treatments has become highly plausible. It has long been known that birds are clever and flexible in their behaviour. But we have apparently seriously underestimated them.

Arie J. van Noordwijk

Arie J. van Noordwijk is at the Netherlands Institute of Ecology, PO Box 40, NL6666ZG Heteren, The Netherlands.
e-mail: a.vannoordwijk@nioo.knaw.nl

(whose transition temperature is 38 K), then we could consider going further, for example to near-linear structures such as LiB_x (ref. 9) at high pressure. Under pressure, boron is also an excellent superconductor in its own right¹⁰.

In a curious historical throwback, a lithium compound was among the class of systems that spawned the idea that electron pairs might be associated with superconductivity. This was Ogg's proposal¹¹, in 1946, based on his studies of metal amines at low temperatures. So perhaps the lithium amine $\text{Li}(\text{NH}_3)_4$ should be squeezed as well. And the possibility of dropping down two further notches in the periodic table, to hydrogen, seems to be back in contention in the quest for higher-temperature superconductors¹², as Shimizu *et al.* cogently observe. ■

N. W. Ashcroft is at the Laboratory of Atomic and

Solid State Physics, Clark Hall, Cornell University, Ithaca, New York 14853-2501, USA.

e-mail: nwa@ccmr.cornell.edu

1. Shimizu, K., Ishikawa, H., Takao, D., Yagi, T. & Amaya, K. *Nature* 419, 597–599 (2002).
2. Matthias, B. T. *Phys. Rev.* 97, 74–76 (1955).
3. Struzhkin, V. V., Hemley, R. J., Mao, H.-K. & Timofeev, Y. A. *Nature* 390, 382–384 (1997).
4. Lin, T. H. & Dunn, K. J. *Phys. Rev. B* 33, 807–811 (1986).
5. Hanfland, M., Syassen, K., Christensen, N. E. & Novikov, D. L. *Nature* 408, 174–178 (2000).
6. Neaton, J. B. & Ashcroft, N. W. *Nature* 400, 141–144 (1999).
7. Christensen, N. E. & Novikov, D. L. *Phys. Rev. Lett.* 86, 1861–1864 (2001).
8. Ginzburg, V. L. & Kirzhnits, D. A. *High-Temperature Superconductivity* (Consultants Bureau, New York, 1982).
9. Wörle, M. & Nesper, R. *Angew. Chem.* 112, 2439–2443 (2000).
10. Eremets, M. I. *et al.* *Science* 293, 272–274 (2001).
11. Ogg, R. A. *Phys. Rev.* 69, 243–244 (1946).
12. Ashcroft, N. W. *Phys. Rev. Lett.* 21, 1748–1749 (1968).
13. Hemley, R. J. & Mao, H.-K. in *Proceedings of the International School of Physics 'Enrico Fermi': High Pressure Phenomena* (eds Hemley, R. J., Bernasconi, M., Ulivi, L. & Chiarotti, G.) (Italian Phys. Soc., Bologna, in the press).

Cancer

The silence of the genes

Bruce R. Zetter and Jacqueline Banyard

The discovery that a transcriptional repressor is turned on in prostate tumours as they become metastatic, leading to the silencing of many genes, suggests a new mechanism for tumour progression.

Tumour cells don't usually start out bad. They generally acquire their malignant properties over years or decades, going from benign to invasive to metastatic to lethal, in a process known as tumour progression¹. This process occurs through a series of mutations that result in the accumulation of aggressive characteristics by the tumour cells². There is also evidence that large numbers of cancer-related genes can simply be switched on or off over relatively short periods of time. This suggests the existence of control molecules that can alter the expression of large groups of genes, thereby influencing tumour progression. On page 624 of this issue, Varambally and colleagues³ describe such a molecule, which may silence a whole constellation of genes during the progression of human prostate cancer.

Previously, this same group of researchers used DNA microarrays, also known as gene chips, to compare genes expressed in normal prostate glands with those in early and advanced prostate cancers⁴. Curiously, their results showed that far more genes were repressed than activated (or 'transcribed') as the prostate tumours became metastatic — that is, acquired the ability to spread to distant organs.

Varambally *et al.*³ now show that a major gene that is activated in advanced prostate cancer encodes the EZH2 protein, a known repressor of gene transcription⁵. They report that when the EZH2 gene is activated in prostate-cancer cells, a substantial number

of other genes are shut down. If some of the products of these genes suppress tumour development, their repression by EZH2 could accelerate a tumour's progress towards metastasis (Fig. 1).

Varambally *et al.* first found that the level of messenger RNA encoding EZH2 significantly increased (indicating increased transcription of the EZH2 gene) in malignant versus benign human prostate cancers. This finding is supported by a recent report⁶ of a 12-fold increase in EZH2 mRNA levels in prostate-tumour metastases compared with

tumours still confined to the prostate. Realizing that they had uncovered a tumour-related gene repressor, Varambally *et al.* then increased EZH2 levels in prostate-cancer cells in culture, and examined the consequences.

They found that the expression levels of 163 genes fell, whereas no genes appeared to be activated — consistent with a functional role for EZH2 as a transcriptional repressor in tumour cells. The authors also conducted a complementary experiment, using small interfering RNA to repress the EZH2 gene in cultured prostate-cancer cells. This resulted in a dramatic decrease in the rate of cell proliferation, showing a direct effect of EZH2 on a specific tumour-promoting function. These results imply that the up-regulation of EZH2 in prostate-cancer cells leads to transcriptional repression and increased cell proliferation. Moreover, the mechanism may not be limited to prostate cancer, as EZH2 expression is also increased in certain lymphomas⁷.

Do the genes that are repressed by EZH2 have the potential to inhibit tumour progression? In some cases, the answer is yes. For example, a number of the EZH2-repressed genes encode putative tumour suppressors, such as the transcription factor *Znfn1a1*, or proteins that inactivate cell signalling, such as Rho GTPase-activating protein 1. On the other hand, some of the changes are counter-intuitive, involving genes whose functions one might expect to see upregulated as cancer progresses. These genes encode, for instance, MMP-7, an enzyme associated with the ability of tumour cells to invade surrounding tissues; and v-erb-b2/HER2, a growth-factor receptor associated with increased cell proliferation. It is important to remember that, although molecules such as EZH2 may lead to gene repression in cancer cells, gene activation also occurs during tumour

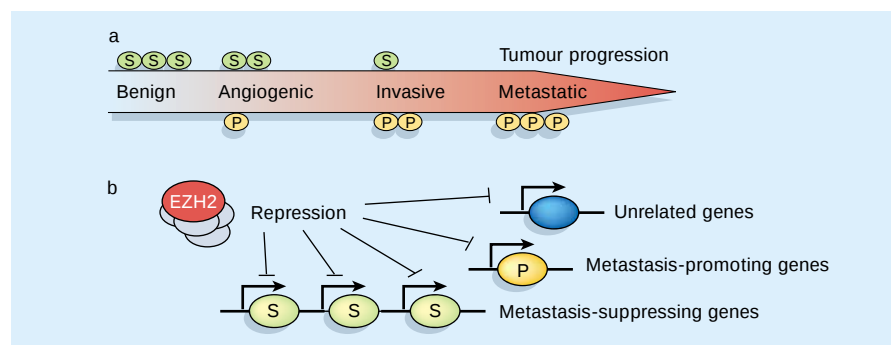


Figure 1 Tumour progression and transcriptional repressor proteins. a, Tumour cells progress from benign, to angiogenic (meaning that they can recruit their own blood supply), to invasive (they can move through surrounding tissue), to metastatic (they can spread to other organs). At the genetic level, this progression is characterized by the net loss of functional metastasis-suppressor genes (S), together with the net gain of active tumour- or metastasis-promoting (P) genes^{1,2}. b, Varambally *et al.*³ have found that the EZH2 protein is activated as prostate tumours progress towards metastasis, and that this protein represses the transcription of whole groups of other genes, many of which may inhibit tumour progression.

progression², and these processes are not mutually exclusive. Over time, it is the balance between enhancers and suppressors of metastasis that will determine the rate of tumour progression (see Fig. 1).

Microarray technology has revolutionized our ability to detect the temporal molecular changes that occur as cancers progress towards metastasis. Combining this technology with advances in bioinformatics provides growing sophistication for analysing such molecular patterns. What is commendable about Varambally and colleagues' approach³ is that, in addition to conducting complex pattern analysis, they have gone on to investigate the potential functions of individual genes. In the study described here, they selected one promising candidate for more intensive investigation using the tools of contemporary molecular and cellular biology. Their results provide new insight into the role of one particular protein, EZH2, in tumour biology, and in so doing reveal a potential new mechanism underlying tumour progression.

Could prostate-cancer patients benefit from this finding? Quite possibly. Although blood levels of the prostate-specific antigen (PSA) protein allow physicians to diagnose prostate cancer, this test is less useful for

prognosis. What is needed is a way of distinguishing prostate cancers that are, or will become, metastatic from those that will remain localized to the prostate. The best prognostic indicators would be molecules whose levels fall or rise during the transition to metastasis—as is the case with EZH2. Significantly, Varambally and colleagues' data³ also indicate that the presence of EZH2 at the time of diagnosis correlates with future tumour recurrence. So this protein, along with other molecules such as thymosin beta 15 (ref. 8) and PTEN⁹, may offer physicians new predictive tools with which to guide decisions about how and when to treat prostate-cancer patients. ■

Bruce R. Zetter and Jacqueline Banyard are in the Department of Surgical Research, Enders 1074, 300 Longwood Avenue, Children's Hospital, Boston, Massachusetts 02115, USA.

e-mails: bruce.zetter@tch.harvard.edu

jacqueline.banyard@tch.harvard.edu

1. Nowell, P. C. *Science* **194**, 23–28 (1976).
2. Hanahan, D. & Weinberg, R. A. *Cell* **100**, 57–70 (2000).
3. Varambally, S. et al. *Nature* **419**, 624–629 (2002).
4. Dhanasekaran, S. M. et al. *Nature* **412**, 822–826 (2001).
5. Laible, G. et al. *EMBO J.* **16**, 3219–3232 (1997).
6. LaTulippe, E. et al. *Cancer Res.* **62**, 4499–4506 (2002).
7. van Kemenade, F. J. et al. *Blood* **97**, 3896–3901 (2001).
8. Chakravarti, A. et al. *Urology* **55**, 635–638 (2000).
9. McMenamin, M. E. et al. *Cancer Res.* **59**, 4291–4296 (1999).

Nanotechnology

Beyond the silicon roadmap

Neil Mathur

Silicon will eventually fail to satisfy the 'smaller, faster, cheaper' drive in technology. Nanoscale techniques could take over, and a recent conference reviewed the prospects for computing and electronics.

When and how will nanotechnology make computers work better? At a conference held last month*, Jim Hutchby (Semiconductor Research Co.) predicted that nanoscale electronics would appear on the International Technology Roadmap for Semiconductors¹ in 2011–2016—and would tear it up by 2050.

Existing silicon-based CMOS technology in computers cannot navigate the roadmap as far as the nanoscale. At that scale, quantum effects start to become significant, and even if silicon components were shrunk to these dimensions the result would not be a 'quantum computer': heat dissipation (around 100 W cm⁻², according to L. Manchanda, Semiconductor Research Co.) would not only lead to the decoherence of fragile quantum states, it would melt the silicon substrate. So what does the science of today suggest about the computing of tomorrow?

In the short term, the existing von

Neumann architecture of today's computers should survive and carry silicon a little beyond its natural shelf-life, thanks to spectacular advances in molecular electronics. For example, binary information can be written into an 'atomic switch' between crossed and nearby nanometre-size wires of platinum and silver if the latter are coated with the solid electrolyte Ag₂S (M. Aono, RIKEN). The silver in this material is mobile², and a few atoms can be electrically induced to migrate to the platinum wire and close the switch. It is possible to combine switches such as these to form logic gates,

the building-blocks of computers, and to perform computations.

Alternatively, rotaxane molecules can be used as binary memory bits. Rotaxanes have a molecular ring surrounding a central linear molecule, like a bead on a wire. The molecular ring can be repositioned by an applied voltage³, altering the rotaxane's longitudinal electrical conductivity. To address a nanoscale array of these molecules using platinum wires requires a clever patterning trick. One possibility is to differentially etch the edge of a GaAs/AlGaAs superlattice, deposit platinum, and transfer the platinum nanowires thus formed to another substrate (J. Heath, UCLA). A 64-bit rotaxane-based memory device produced within one square micrometre has now been demonstrated (S. Williams, Hewlett–Packard). This level of bit density lies quite a few years down the silicon roadmap, and could in fact be increased by a few orders of magnitude by incorporating patterning tricks such as the one described above.

Various groups have demonstrated that single-wall carbon nanotubes can also be used to make an impressive variety of room-temperature electronic logic devices (for example, refs 4 and 5). In contrast, DNA has been found to be an unsuitable building-block for computers because it is electrically insulating (C. Dekker, Delft Univ.). Although basic logic functions have now been performed in a range of nanostructures, fault-free mass fabrication is, at least for now, thought to be impossible. So computers that incorporate nanotechnology in the near future will have to have some degree of fault tolerance, and for this, novel architectures are required. One strategy, used in the Hewlett–Packard 'Teramac'⁶, is to build in a slight redundancy, so that perfect performance can be achieved provided there are sufficiently few faults⁷.

But what about more radical departures from the conventional architecture? Perhaps phase-based logic will overtake amplitude-based logic (L. Manchanda). This leads us to consider quantum computing. A quantum computer with just a few bits has already been produced experimentally using nuclear magnetic resonance and a solution of organic molecules⁸. Just one of many recent ideas (J. C. Egues, Basel Univ.) is based on controlling electron spin using gate-induced spin-orbit coupling⁹. In fact,



Figure 1 The conference logo writ small. Carbon nanotubes such as these could make both the computers and displays of tomorrow. (Image courtesy Kenneth Teo, Manish Chhowalla, Gehan Amarutunga and Bill Milne, Univ. Cambridge.)

*Trends in Nanotechnology, Santiago de Compostela, Spain, 9–13 September 2002.

many possible routes to quantum computing have been suggested, but the most promising are solid-state implementations — most famously, nuclear spins of phosphorus atoms in a silicon matrix¹⁰ — because they can be scaled up to generate the massive parallelism required for useful computation.

The counter-intuitive rules of quantum mechanics imply that, unlike classical computers, quantum computers should perform best with a slow clock speed¹¹ (in other words, the devices within them should switch slowly). This necessarily suggests that rapid decoherence of the quantum states encoding information is unacceptable. Entanglement between two solid-state quantum bits, or qubits, was reported for the first time at this conference — albeit with a rather rapid decoherence time of 300 picoseconds (J. S. Tsai, NEC). The qubits used were based on low-temperature superconducting tunnel junctions. Perhaps this breakthrough indicates where the future of quantum computing actually lies.

Another departure from conventional computing philosophy would be to work with light rather than electrons. Whereas the periodicity of the structure of crystalline materials is comparable to the wavelength of mobile electrons, advanced materials techniques enable periodic structures to be produced in which the repeating unit matches the wavelength of light — these are photonic band-gap materials. Complex structures formed with these advanced techniques (for example, ref. 12) could be used to manipulate light and form computer logic gates. Indeed, computer simulations suggest that light can bend around cleverly constructed corners with no discernible energy loss (S. John, Univ. Toronto).

The conference was also presented with some aspects of nanotechnology that could have an indirect impact on tomorrow's computers. For example, the issue of heat dissipation mentioned earlier could be locally monitored using a 75-nm-diameter carbon nanotube containing the liquid-metal gallium, which would act like a nanoscale mercury thermometer¹³ in the 50–500 °C range (Y. Bando, National Institute for Materials Science, Tsukuba). Or perhaps computerized devices such as gas-specific sensors will rely on specially adapted nanotube elements — such devices can detect as little as 0.02% H₂ (G. Gruner, Nanomix, Inc. and UCLA). Or perhaps computer display devices will use electron emission from nanotubes (P. Legagneux, Thales) such as those shown in Fig. 1 — in fact, Samsung have this year reported a proof-of-principle display unit¹⁴.

So where will it all end? The graphics of S. John were reminiscent of the best that Hollywood can achieve, and the many talks relating to biology could at times make this

physicist feel that he had come to the wrong conference. The nanoscale therefore seems to be the length scale at which science has become truly interdisciplinary. The question is, will it reconcile science fact with science fiction? ■

Neil Mathur is in the Department of Materials Science and Metallurgy, University of Cambridge, New Museums Site, Pembroke Street, Cambridge CB2 3QZ, UK.
e-mail: ndm12@cus.cam.ac.uk

1. International Technology Roadmap for Semiconductors (2001); <http://public.itrs.net>
2. Terabe, K., Nakayama, T., Hasegawa, T. & Aono, M. *J. Appl. Phys.* **91**, 10110–10114 (2002).

3. Collier, C. P. *et al. Science* **285**, 391–394 (1999).
4. Bachtold, A., Hadley, P., Nakanishi, T. & Dekker, C. *Science* **294**, 1317–1320 (2001).
5. Huang, Y. *et al. Science* **294**, 1313–1317 (2001).
6. Heath, J. R., Kuekes, P. J., Snider, G. S. & Williams, R. S. *Science* **280**, 1716–1721 (1998).
7. Kuekes, P. J. & Williams, R. S. Demultiplexer for a molecular wire crossbar network. US patent 6,256,767 (2001).
8. Jones, J. A., Mosca, M. & Hansen, R. H. *Nature* **393**, 344–346 (1998).
9. Datta, S. & Das, B. *Appl. Phys. Lett.* **56**, 665–667 (1990).
10. Kane, B. E. *Nature* **393**, 133–137 (1998).
11. Feynman, R. P. *Feynman Lectures on Computation* (eds Hey, A. J. G. & Allen, R. W.) (Penguin, London, 1999).
12. Toader, O. & John, S. *Science* **292**, 1133–1135 (2001).
13. Gao, Y. H. & Bando, Y. *Nature* **415**, 599 (2002).
14. Chung, D. S. *Appl. Phys. Lett.* **80**, 4045–4047 (2002).

Ecology

Biodiversity in the scales

Nicholas J. Gotelli

Scaling coefficients can reveal nonlinear relationships among biological variables. The approach has now proved fruitful in exploring the relationship between diversity at different taxonomic levels.

Species diversity is central to ecological studies. But it is a challenging parameter to measure because diversity is organized hierarchically: individual organisms are classified into species, species into genera, genera into families, and so on. As a consequence, patterns of diversity at one level (genera, for instance) are linked statistically and by evolutionary history to the patterns at higher and lower levels (such as families and species). On page 610 of this issue¹, Enquist and colleagues describe tight scaling relationships between species richness and diversity among the higher taxa (genera or families), in both living and fossil plant communities. These patterns should eventually help in understanding how species diversity is controlled and how total biomass is partitioned among coexisting species.

Enquist *et al.*¹ used allometric scaling equations that have traditionally been applied to problems of body size and relative growth. In mammals, for example, a power function² describes the relationship between brain size (x) and body size (y):

$$\text{Brain size} \propto \text{constant} \times (\text{body size})^z$$

The exponent z is the scaling coefficient, and describes whether y is increasing faster ($z > 1$) or more slowly ($z < 1$) than x . On a log–log scale, these relationships plot as straight lines for which the scaling coefficient is the slope. In previous work, Enquist and colleagues^{3–5} successfully applied allometric models to general patterns of biomass, body size and growth form in plant and animal assemblages. Here, they extend this approach to describe patterns of taxonomic and biomass partitioning in different tree

communities. They analysed a high-quality global data set, collected by the late Alwyn H. Gentry, that includes 227 sites, each one-tenth of a hectare in area, from six continents. A database of plant fossil communities from the Tertiary period included 29 samples from sites in western North America.

For both the living and fossil tree communities, Enquist *et al.* plotted the logarithm of the number of higher taxa against the logarithm of the number of species, and discovered a tight relationship that explained more than 90% of the variance (although the fit to the Gentry data was not strictly linear). In other words, the number of higher taxa in a community is highly predictable from the number of lower taxa. In one sense, this is not new. Ecologists and biogeographers established long ago that the diversity of higher taxa changed predictably in large communities compared with small ones⁶. As more species are drawn randomly from a regional source pool, the number of higher taxa rises steeply at first as common taxa are added, and then more slowly as progressively rarer taxa are added⁷. The power function transforms this sampling curve into an approximately linear allometric relationship over limited regions of the curve.

Could taxonomic scaling be the inevitable consequence of random sampling from a large regional or global species pool? The answer seems to be no: the exponents for randomly constructed communities were consistently larger than those found by Enquist and colleagues. This result means that higher-level diversity in communities accumulates relatively slowly. Put another way, communities consist of slightly more species per genus or species per family than

would be expected by chance, a pattern first noted in the 1940s by the statistical ecologist C. B. Williams⁸.

If statistical sampling cannot entirely account for the pattern, what is the mechanism controlling taxonomic diversity? Previous applications of allometric scaling laws have been successful because they have often been derived from first principles of physical scaling laws, such as the hydrodynamic constraints that control the flow of fluids through plants⁵. Here, there are no such underlying first principles, and the mechanisms controlling taxonomic diversity are elusive.

Closely related species have similar dispersal abilities, so that communities might come to be dominated by families or genera that contain good dispersers, which would decrease the scaling coefficient. However, Enquist and colleagues' null model assumed that species colonize randomly and with equal probability, so it did not incorporate differences in dispersal ability among species or differences in total abundance among sites, which ranged from 52 to 1,005 individuals. Moreover, the fixed plot sizes in the Gentry data set will have sampled different proportions of species-poor and species-rich communities. As a consequence, the allometric coefficients measured by Enquist *et al.*

are not constants and will change with plot size or sampling design. Variations in annual precipitation, site elevation and community age also contributed to deviations from the null-model predictions. Enquist *et al.* propose that local and regional processes control the pattern, but much more analysis will be necessary to identify the specific mechanisms responsible for this tight association.

Taxonomic scaling may influence other patterns. In previous analyses of the Gentry data, Enquist and Niklas⁹ established that total biomass per plot varied surprisingly little across different communities. Consequently, as more species are added to a community, the amount of biomass partitioned to each species must inevitably decrease. This result appears to contradict studies in which total biomass increases when species are experimentally added to communities¹⁰. But there is enough scatter in the biomass partitioning data to accommodate both patterns. In contrast to the tight taxonomic scaling curves, the allometric relationship between biomass per taxon and total species richness explained little more than 50% of the variance in the data. For a given number of species in a community, biomass per species can vary tenfold. At smaller spatial scales, adding species to a community may indeed increase total bio-

mass, even though biomass per species falls when communities are compared across large biogeographical regions.

The application of allometric scaling laws to patterns of taxonomic diversity has revealed surprising regularities in the diversity of plant communities. Moreover, the existence of such patterns in both fossil and living plant assemblages from different regions suggests that diversity might be regulated more by local processes, such as species interactions, than by historical factors, such as dispersal barriers or speciation. Teasing apart the specific biological and statistical mechanisms responsible for these patterns is a task for the future.

Nicholas J. Gotelli is in the Department of Biology, University of Vermont, Burlington, Vermont 05405, USA.

e-mail: ngotelli@zoo.uvm.edu

1. Enquist, B. J., Haskell, J. P. & Tiffney, B. H. *Nature* **419**, 610–613 (2002).
2. Pagel, M. D. & Harvey, P. H. *Science* **244**, 1589–1593 (1989).
3. West, G. B., Brown, J. H. & Enquist, B. J. *Science* **276**, 122–126 (1997).
4. Enquist, B. J., Brown, J. H. & West, G. B. *Nature* **395**, 163–165 (1998).
5. Enquist, B. J. & Niklas, K. J. *Science* **295**, 1517–1520 (2002).
6. Järvinen, O. J. *Biogeogr.* **9**, 363–370 (1982).
7. Gotelli, N. J. & Colwell, R. K. *Ecol. Lett.* **4**, 379–391 (2001).
8. Williams, C. B. *Patterns in the Balance of Nature* (Academic, New York, 1964).
9. Enquist, B. J. & Niklas, K. J. *Nature* **401**, 655–660 (2001).
10. Loreau, M. *et al.* *Science* **294**, 804–808 (2001).

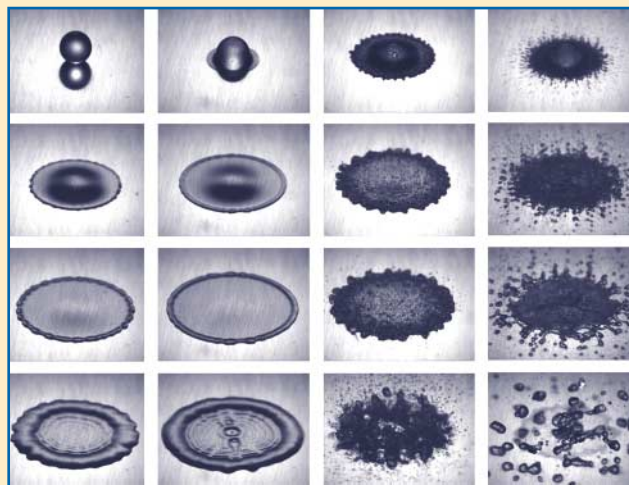
Fluid mechanics

Impact factors

Fire extinguishers, such as sprinkler systems, typically deliver a fine, pressurized spray of water droplets. Some of the smaller water droplets from a sprinkler may be deflected by the plume of the fire and so do little to extinguish it directly. But hot surfaces around the fire can be cooled by the evaporation of droplets that land on them, and the resulting vapour may also help to stifle the flames.

Because of environmental concerns, the widely used fire-quenching agent Halon 1301 was banned by international treaty in 1987. An alternative is to use water with some additive, such as sodium acetate trihydrate, that enhances its fire-quenching ability. How does that affect droplet behaviour?

Samuel L. Manzello and Gianni C. Yang (*Proc. R. Soc. Lond. A* **458**, 2417–2444; 2002) have investigated the collision dynamics of pure water droplets and droplets containing 30% (by mass) sodium acetate trihydrate as they impact



on a hot metal surface. Whether each droplet spreads, splashes or rebounds depends on its impact energy as well as on the temperature and roughness of the surface.

Manzello and Yang recorded the collisions of 2.7-mm-diameter drops on a stainless-steel surface for a range of surface temperatures and impact energies — the latter

quantified by the 'Weber number', related to the liquid's density and surface tension, and the velocity and diameter of the droplets. The images shown here were taken for Weber number = 181, at times (rows top to bottom) of 0, 1, 2 and 7 ms after collision and at surface temperatures (columns left to right) of 20, 104, 230 and 340 °C.

Compared with the behaviour

of droplets of pure, distilled water, Manzello and Yang found that the presence of sodium acetate trihydrate made quite a difference to the collision dynamics at low impact (low Weber number), but less so for high-velocity impacts. In most fire-extinguishing systems the droplets are delivered from a pressure nozzle at high velocity, so the authors conclude that knowing how pure water droplets behave is a reasonable approximation to the behaviour of water with an additive.

They also note that, for high-impact collisions, a droplet forms a liquid film of greater diameter than in low-impact collisions, which increases its capacity for surface cooling. Although the relation of liquid-film diameter to surface temperature was clear for droplets of pure water, Manzello and Yang struggled to find such a correlation for water droplets containing the additive. More work, both theoretical and experimental, is needed here.

Alison Wright

Quantum physics

Single photons stick together

Philippe Grangier

In the right circumstances, two photons can meet and 'coalesce'. This effect has now been observed for photons emitted independently from a single-photon source, and has implications for quantum computing.

Can two photons that have never met know something about each other? The question bothered Einstein, and, reporting on page 594 of this issue, Santori *et al.*¹ demonstrate in a new context that the answer is 'yes'. They have found that consecutive photons emitted from a single-photon source may indeed know each other sufficiently well to be able to interfere in a very peculiar way.

From the perspective of quantum physics, the answer to the above question is 'yes indeed, provided that the photons are in the same, single mode'. Photons are characterized by several physical properties, such as their frequency and polarization, and together these properties define the 'mode' of the electromagnetic field with which the photon is associated. In the simplest approach, a mode can be regarded as a polarized, monochromatic, plane wave, but it can also be defined in a much more flexible way: the photon can be spread coherently over several modes, defining a new single mode.

If two photons are in the same mode, quantum mechanics predicts that a bunching, or coalescence, effect occurs. Let's consider two photons sent into a beam splitter where each has a 50–50 chance of being reflected or transmitted. Let's also assume that one photon that is being transmitted ends up in exactly the same mode as the other photon being reflected — in other words, all the properties of the two photons are identical at the beam-splitter output, so that they become essentially indistinguishable. In that case, there are four possible configurations for the two photons being transmitted or reflected, as depicted in Fig. 1.

As is usual in quantum mechanics, there is a probability amplitude for each of these configurations, and the transmission and reflection probabilities are given by the square of the modulus of the sum of these amplitudes. According to this simple calculation, two amplitudes (Fig. 1b, c) have opposite signs, and so cancel each other — they interfere destructively. The result is that the two photons must go to the same output beam (Fig. 1a, d): it is as though they coalesce as they meet at the beam splitter.

This surprising quantum-interference effect, which can be attributed to the 'bosonic' character of photons, was first predicted and observed by Hong, Ou and Mandel² in 1987. They used pairs of 'twin photons', produced simultaneously in a process called parametric down-conversion, which can be

induced by shining laser light into a crystal. An optical delay was added to the path of one of the two photons. If this delay is not zero, each one of the twin photons is randomly transmitted or reflected — they miss each other. But if the delay is exactly zero, the photons overlap and the coalescence effect occurs.

Experimentally, it is easy to count the number of times that one photon is transmitted and the other is reflected. As the value of the optical delay is scanned through zero, the number of transmitted–reflected photon pairs vanishes, creating the so-called Mandel dip that is striking evidence of the coalescence effect. It could be argued that the two photons knew about each other before entering the beam splitter because they were twins emitted in a single fluorescence event. But Santori *et al.*¹ have demonstrated that the same effect occurs for truly independent photons, emitted one at a time from a semiconductor quantum dot.

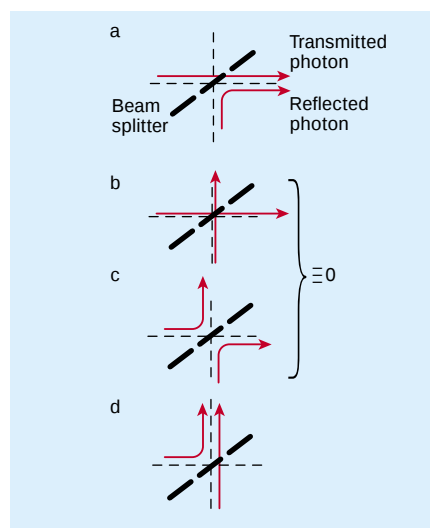


Figure 1 Two of a kind. When two single-mode, but otherwise independent, photons enter a 50–50 beam splitter, they may be transmitted or reflected in various ways: one may be reflected and the other transmitted (a and d); both photons may be transmitted (b), or both may be reflected (c). By quantum interference, the probabilities for 'both transmitted' and 'both reflected' cancel out. Only outcomes a and d are possible, and the two photons emerge from the beam splitter along the same output, as though they had coalesced. Santori *et al.*¹ show that this quantum effect holds for independent photons emitted from a single-photon source.

Emission of indistinguishable single photons from quantum dots³ (or single atoms⁴) can be engineered using cavity quantum-electrodynamics effects⁵. Santori *et al.*¹ sandwiched semiconductor quantum dots between mirrors to form a microcavity and then varied the delay between consecutive emitted photons arriving at a beam splitter. When the photons arrived at the beam splitter together, Santori and colleagues saw the distinctive Mandel dip caused by coalescence of the photons.

This coalescence is not just a curious quantum effect, but has fundamental implications in the field of quantum information processing. Proposed all-optical quantum-photonics networks^{6,7} are based on indistinguishable single photons carrying information from node to node and interacting with one another. For such schemes to work, these photons should be in a single mode, exactly in the sense explained above.

Spectacular progress has been made with single-photon sources, and it is now possible to produce single photons on demand, and to use them in quantum-cryptography systems⁸. But most sources produce photons that are spread incoherently over many modes, preventing adequate single-mode behaviour. Parametrically generated photon pairs can be used in some circumstances, such as for teleportation⁹, but these are also not ideal as they are not emitted on demand, and thus require an extra unwanted step of conditional preparation.

Santori and colleagues' experiment, demonstrating the indistinguishability of photons from a single-photon source, can be seen as a first step towards making quantum logic gates for photon-based quantum computing⁶. But the difficulties should not be underestimated: the error rates generated using the present set-up would be too large (by orders of magnitude) compared with the range over which quantum error-correcting codes could be effective. Also, the number of interfering photons required to implement a useful computation is huge, and the integration of the devices would have to be pushed far beyond present technological abilities. Nevertheless, this experiment is a remarkable demonstration of the very peculiar properties of the 'purest' states of light that can be imagined — single photons in single modes.

Philippe Grangier is at the Institut d'Optique, F-91403 Orsay, France.

e-mail: philippe.grangier@iota.u-psud.fr

1. Santori, C., Fattal, D., Vučković, J., Solomon, G. S. & Yamamoto, Y. *Nature* **419**, 594–597 (2002).
2. Hong, C. K., Ou, Z. Y. & Mandel, L. *Phys. Rev. Lett.* **59**, 2044–2047 (1987).
3. Moreau, E. *et al.* *Appl. Phys. Lett.* **79**, 2865–2867 (2001).
4. Kuhn, A., Hennrich, M. & Rempe, G. *Phys. Rev. Lett.* **89**, 067901 (2002).
5. Gérard, J. M. *et al.* *Phys. Rev. Lett.* **81**, 1110–1113 (1998).
6. Knill, E., Laflamme, R. & Milburn, G. J. *Nature* **409**, 46–52 (2001).
7. Cirac, J. I., Zoller, P., Kimble, H. J. & Mabuchi, H. *Phys. Rev. Lett.* **78**, 3221–3224 (1997).
8. Beveratos, A. *et al.* Preprint quant-ph/0206136 (2002); <http://arXiv.org>
9. Bouwmeester, D. *et al.* *Nature* **390**, 575–579 (1997).

Obituary

George Porter (1920–2002)

The time taken for molecules to dissociate or rearrange lies typically in the range 1–100 fs, a femtosecond being 10^{-15} s. When, in 1945, George Porter started out as a research student at the University of Cambridge, direct observation of these processes was impossible. But he lived to see the techniques that he developed reach the astonishing state today where such ultrafast events can be studied routinely in real time. Porter died on 31 August 2002, aged 81. He was scientifically active until the last two years of his life, a tribute to his devotion to his own and his colleagues' science, and to promoting science generally.

Porter, a Yorkshireman, displayed a passion for science as a schoolboy and obtained his BSc in chemistry at the University of Leeds. At Cambridge, with R. G. W. Norrish, he tackled the problem of how to detect short-lived chemical intermediates, so-called free radicals. These are molecules that have an unpaired electron, which makes them extremely chemically reactive, so that they exist for what was thought then to be a very short duration — typically milliseconds or less. Porter knew that free radicals could be produced easily by absorption of light. From his wartime experience with signalling lamps using intense pulses, he reasoned that if a radical were created by a pulse of light that was short in comparison to the radical's lifetime, the species could be detected absorbing light from either a second continuous light source or a second pulse of light delayed in time with respect to the first. These experiments were spectacularly successful. They led ultimately to Porter and Norrish's award of a share in the 1967 Nobel Prize in Chemistry, for their development of 'flash photolysis'.

Radicals are common intermediates in chemistry and biology, and so of great interest, but there are even shorter-lived chemical entities, such as electronically excited states. Porter turned flash photolysis to the task of observing intermediates such as 'triplet states' in the microsecond time-domain, and by 1960 this had become a standard approach. However, for conventional light sources, the shorter the pulse, the weaker the light. Nanosecond pulses can be made, but the intensity is too small — about as bright as a spark plug — to be useful.

In 1963, Porter had moved from Cambridge to the University of Sheffield,



Innovator in ultrafast chemistry and advocate for pure research

where he was Firth professor of chemistry (he joked that with more funds, they would have appointed a 'Thecond' professor). Three years later, he and his group transferred to the Royal Institution in London — the RI. There they took advantage of Theodore Maiman's invention of the visible ruby laser, which could produce both short pulses and high intensities. Porter and colleagues exploited the laser technique, first in the nanosecond region, then quickly to picoseconds and, after 1985, at Imperial College London, ultimately to the femtosecond regime. He relished pointing out that within his lifetime the time-domain accessible to the 'ultrafast scientist' had shortened by about the same number of orders of magnitude as the timescale from the beginning of the Universe to the present day.

Few chemical reactions initiated by light are as important as photosynthesis, by which sunlight, carbon dioxide and water produce the carbohydrate food source that underpins all animal life on the planet. Porter's group helped to lay the foundations in understanding this vital process by investigating the photo-induced transfer of electrons in the first few picoseconds of the reactions. Inspired by this application of light in nature, he

became increasingly interested in the artificial use of sunlight to provide energy, either directly as electricity, or by splitting water to produce hydrogen as a clean fuel.

As director of the RI, George Porter was inspirational. No one could forget the Davy–Faraday Laboratory weekly discussions — George sitting in the front row, quietly puffing on his pipe under the no-smoking sign, and listening to young students or distinguished visitors with equal courtesy, and through politely expressed questions cutting straight to the heart of the matter (often to the discomfort of the speaker). The RI was founded in 1799, and by 1966 could easily have been regarded as an anachronism. But in the 19 years they were in residence, Porter and his wife Stella made it a living centre for the exposition of science. They brought a memorable style and glamour to the public functions; Davy and Faraday would have approved of the preservation of top-class science and lucid scientific presentation.

Popularization of science was then widely sneered at, but Porter had the courage and vision to make it a major concern and changed the perception of such activities. He insisted that speakers at the RI discourses use demonstrations in the best traditions of the institution, and exemplified the style by being the most polished of presenters. His early films on thermodynamics still look remarkable. At the RI, he also introduced the Christmas Lectures to BBC television audiences. Uniquely, he simultaneously held the positions of president of the Royal Society, president of the British Association for the Advancement of Science, and director of the Royal Institution, and he was instrumental in the founding of COPUS — the Committee for the Public Understanding of Science.

George Porter was a powerful advocate of preserving a strong scientific base in Britain, saying "pure research was merely that research which has not yet been applied". He had the ear of senior politicians, and after his creation as Lord Porter of Luddenham in 1990 he vigorously promoted the cause of pure research in the House of Lords.

He will be badly missed by his colleagues and collaborators, who number hundreds, and by his family — particularly Stella, who supported him with a passion equal to his own for science. David Phillips
David Phillips is in the Department of Physical Chemistry, Imperial College, Exhibition Road, London SW7 2AZ, UK.
e-mail: d.phillips@ic.ac.uk

Buckling cascades in free sheets

Wavy leaves may not depend only on their genes to make their edges crinkle.

The edge of a torn plastic sheet forms a complex three-dimensional fractal shape. We have found that the shape results from a simple elongation of the sheet in the direction along its edge. Natural growth processes in some leaves, flowers and vesicles could lead to a similar elongation and hence to the generation of characteristic wavy shapes.

We used rectangular plastic sheets pulled from the sides (in the y -direction) to generate a steadily travelling crack (in the x -direction). The high stresses near the crack tip produce an irreversible plastic deformation of the sheet and, as they are relieved, the deformed sheet is free to relax and to adopt a new shape in space.

Surprisingly, the equilibrium shape of the sheet consists of a cascade of waves upon waves along the newly formed edge. The waveform along the edge examined at six levels of magnification is self-similar, with a scaling factor of 3.2 (Fig. 1a). The amplitudes, A , of the waves in the cascades are simply related to their wavelengths λ by $A = 0.15\lambda$ (Fig. 1b); the edge of the sheet is therefore a fractal. The fractal scaling spans 2.5 orders of magnitude and stops at a small length scale that is 6.5 times the sheet's thickness. Measurements of the deformation field revealed an increase in the length of the sheet in the x -direction that depends only on y , and decays smoothly to zero over a distance of a few millimetres from the edge (Fig. 1c).

How can such featureless deformations lead to such complex shapes? The key to this is to realize that the bending rigidity of thin plates is much smaller than their stretching rigidity¹. Thus, to reduce its elastic energy, the sheet can easily buckle into shapes that remove in-plane compression. The thinner the sheet and the larger the compression, the smaller are the possible wavelengths of the buckles. This principle is seen in action during the crumpling of paper sheets^{2–5} and the blistering of thin films^{6,7}, where sheets adopt very complex shapes in response to a uniform external loading.

Unlike these examples, our sheets (after tearing is complete) are free and not subjected to any external loading. The spontaneous buckling occurs as a consequence of the stresses created by the differential deformation within the sheet. Far from the edge, the deformation is small (Fig. 1c), and the compression leads to buckling that has a single long wavelength. Closer to the edge, the compression increases rapidly, leading to an ordered buckling sequence with smaller and smaller wavelengths.

This phenomenon should be general,

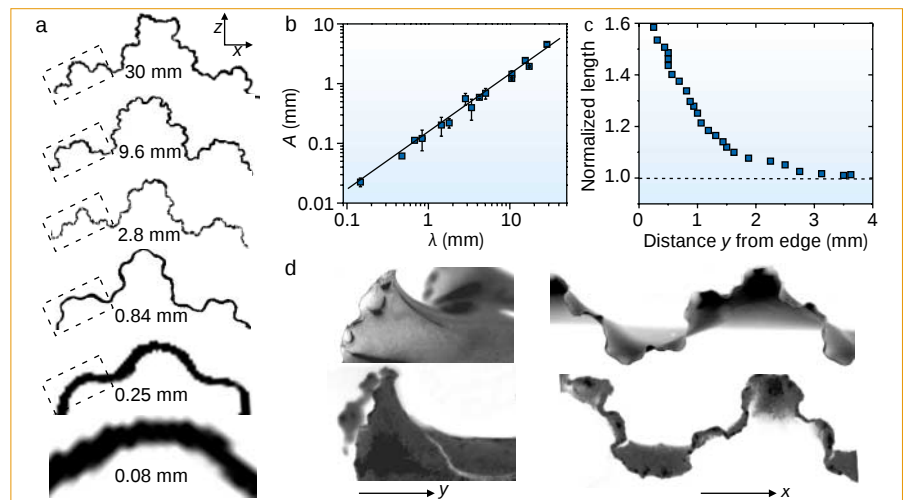


Figure 1 Buckling cascade in a deformed plastic sheet. **a**, Different magnifications of the edge of the sheet (0.012 mm thick). Successive pictures show the dotted boxed region on the left of the previous picture magnified by 3.2; the width of each image is indicated. **b**, The amplitude of a wave in the cascade is related to its wavelength by $A = 0.15\lambda$ (solid line), for sheet thicknesses ranging from 0.012 to 0.5 mm. **c**, The length of a sheet (0.2 mm thick) in the x -direction, normalized by its original length, as a function of distance from the edge. Dashed line indicates the original length. **d**, Side (left) and front (right) views of a plastic sheet (top panels) and a beet leaf (bottom panels), showing the similarity between the two buckled sheets.

and indeed we found buckling cascades in numerical investigations of elastic sheets with deformations similar to those shown in Fig. 1c. Although in the experiment the sheet thinned near the edge (owing to volume conservation), the simulations both with and without thinning revealed buckling cascades. In addition, numerical analysis indicates that the particular scaling properties of the buckling depend on the deformation profile.

A similar increase in length near an edge, and therefore similar internal stresses, can result from growth processes. The wavy structures found in many flower petals, lichens and leaves, such as the beet leaf in Fig. 1d, might be formed in this way. No genetic coding is needed to instruct pieces of a leaf to curl up and curl down. All that is required is a growth process to elongate the sheet along its edge — elasticity takes

care of the rest.

Eran Sharon*, **Benoît Roman***,
Michael Marder*, **Gyu-Seung Shin*†**,
Harry L. Swinney*

**Center for Nonlinear Dynamics and Department of Physics, The University of Texas at Austin, Austin, Texas 78712, USA*

e-mail: eran@chaos.ph.utexas.edu

†Permanent address: Department of Physics, Kyung Hee University, Yongin, Kyunggi, Korea

1. Lord Rayleigh *Theory of Sound* Vol. 1, 396 (Dover, New York, 1945).
2. Ben Amar, M. & Pomeau, Y. *Proc. R. Soc. Lond. A* **453**, 729–755 (1997).
3. DiDonna, B. A., Witten, T. A., Venkataramani, S. C. & Kramer, E. M. *Phys. Rev. E* **65**, 016603-1–016603-25 (2002).
4. Matan, K., Williams, R., Witten, T. & Nagel, S. *Phys. Rev. Lett.* **88**, 076101-1–076101-4 (2002).
5. Kramer, E. & Witten, T. *Phys. Rev. Lett.* **78**, 1303–1306 (1997).
6. Argon, A., Gupta, V., Landis, H. & Cornie, J. *J. Mater. Sci.* **24**, 1207–1218 (1989).
7. Gioia, G. & Ortiz, M. *Adv. Appl. Mech.* **33**, 119–192 (1997).

Competing financial interests: declared none.

Thin films

Wrinkling of an elastic sheet under tension

Here we consider the wrinkling of a stretched, slender elastic sheet and derive scaling laws for the wrinkle wavelength and amplitude that are valid far from the onset of buckling. Our results, which can be generalized to stretched and sheared anisotropic and inelastic sheets,

could form the basis of a sensitive assay for the mechanical characterization of thin solid films.

When such a thin isotropic elastic sheet of thickness t , width W and length L ($t \ll W < L$), composed of a material with Young's modulus E and Poisson's ratio ν , is subjected to a longitudinal stretching strain, γ , in its plane, it remains flat for $\gamma < \gamma_c$, a critical stretching strain. Further stretching causes the sheet to wrinkle (Fig. 1a).

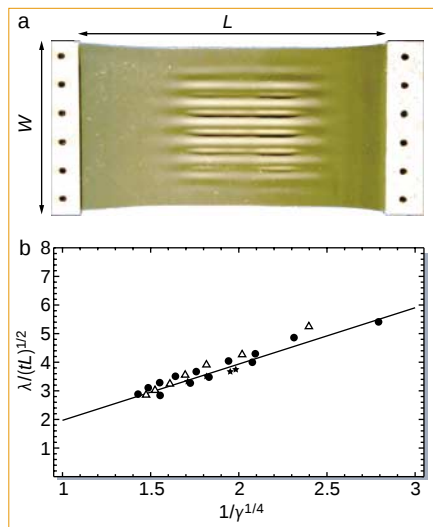


Figure 1 Wrinkles in a polyethylene sheet. **a**, Photograph of a sheet (length, L , 20 cm; width, W , 12 cm; thickness, t , 0.01 cm) under a uniaxial tensile strain, $\gamma \approx 0.1$. **b**, Plot of the dimensionless wavelength $\lambda/(tL)^{1/2}$ against $1/\gamma^{1/4}$ for various lengths of a polyethylene sheet. Lengths: circles, 25.4 cm; stars, 30.4 cm; triangles, 34.5 cm. The data span more than an order of magnitude in strain, with $\gamma \in [0.01, 0.2]$. The line corresponds to the theoretical prediction $\lambda/(tL)^{1/2} = (4\pi^2/3\gamma(1-\nu^2))^{1/4}$. For polyethylene, $\nu \approx 0.35$, so that $\lambda/(tL)^{1/2} \approx 2/\gamma^{1/4}$, which is consistent with the results of our experiments.

This non-intuitive phenomenon occurs because the clamped boundaries prevent the sheet from contracting laterally in their vicinity, setting up a local biaxial state of stress; that is, the sheet is sheared near to its boundaries. The transverse stress is tensile near the clamped boundary and compressive slightly further from it¹. When $\gamma \approx \gamma_c$, the sheet buckles to accommodate the in-plane strain incompatibility generated by the Poisson effect. However, typically $\gamma \gg \gamma_c$, so a linear theory is of little use and we must consider the geometrically non-linear behaviour of the wrinkles.

For a sheet thus stretched, a periodic texture of parallel wrinkles of wavelength $\lambda < W$ decorates most of the sheet. The wavelength and amplitude are determined by minimizing the energy, $U = U_B + U_S$, where U_B is the energy due to bending (primarily in the transverse direction) and U_S is the energy due to stretching (along the wrinkles), subject to any geometric constraints (transverse inextensibility in this case).

If A is the amplitude of the wrinkles, then $U_B \approx Et^3(A/\lambda^2)^2LW$, whereas $U_S \approx Et\gamma(A/L)^2LW$. Transverse inextensibility implies that $(A/\lambda)^2 \approx \nu\gamma$, which quantifies the shrinkage of the sheet induced by the Poisson effect due to the imposed longitudinal strain. From this last relationship, we may derive $U \approx (Et^3/\lambda^2 + Et\gamma\lambda^2/L^2)\nu\gamma LW$. Minimizing U with respect to λ gives a scaling law for the wavelength

$$\lambda \approx (tL)^{1/2}/\gamma^{1/4} \quad (1)$$

Substituting this relationship into the transverse-inextensibility constraint gives a scaling law for the amplitude, $A \approx (\nu tL)^{1/2}\gamma^{1/4}$. A refined calculation (E.C. and L.M., manuscript in preparation) gives the following expressions for the pre-factors: $C_\lambda = \lambda\gamma^{1/4}/(tL)^{1/2} = (4\pi^2/3(1-\nu^2))^{1/4}$, and $C_A = A/(\nu tL)^{1/2}\gamma^{1/4} = (16/3\pi^2(1-\nu^2))^{1/4}$.

To verify our results, we stretched different lengths of a polyethylene sheet of thickness ~ 0.01 cm and width 12 cm, by using extensional strains $\gamma \in [0.01, 0.2]$. The clamped boundary conditions were enforced by using two aluminium plates to sandwich the sheet, attached with double-sided adhesive tape to ensure that no slippage occurred. In Fig. 1b, we plot the experimental values of $\lambda/(tL)^{1/2}$ against $1/\gamma^{1/4}$, together with our theoretical prediction, which shows quantitative agreement.

Our results complement those of classical tension-field theory^{2,3}, which focuses on the much simpler problem of determining the locations of the wrinkles, but not their fine structure. The physics that underpins our calculation is as follows: a geometric packing constraint leads to the formation of wrinkles of intermediate wavelength because the bending resistance of the sheet penalizes short wavelengths and longitudinal stretching penalizes long wavelengths. Indeed, there is a mathematical similarity between the effect of the anisotropic tension on a curved surface and a simple elastic foundation, which allows us to generalize our results to a variety of applications (E.C. and L.M., manuscript in preparation). Moreover, inverting equation (1) yields $\gamma \approx t^2L^2/\lambda^4$, showing that wavelength is an extremely sensitive assay of strain in a wrinkled sheet.

Our results could form the basis of a quantitative wrinkling assay of thin solid films, and can potentially be applied to cell motility⁴, skin biomechanics and the mechanical characterization of soft solid membranes.

E. Cerda*†, K. Ravi-Chandar‡, L. Mahadevan*

*Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge CB3 9EW, UK

e-mail: l.mahadevan@damtp.cam.ac.uk

†Departamento Física, Universidad de Santiago de Chile, Santiago, Chile

‡Department of Aerospace Engineering and Engineering Mechanics, University of Texas, Austin, Texas 78705, USA

1. Benthem, J. Q. *J. Mech. Appl. Math.* **16**, 413–430 (1963).
2. Wagner, H. *Z. Flugtech. Motorluftschiffart* Vol. 20, no. 8–12 (1929).
3. Mansfield, E. *Proc. XII Int. Congr. Theor. Appl. Mech.* (Springer, New York, 1968).
4. Harris, A. K., Wild, P. & Stopak, D. *Science* **208**, 177–179 (1980).

Competing financial interests: declared none.

Anthropogenic aerosols

Indirect warming effect from dispersion forcing

Anthropogenic aerosols enhance cloud reflectivity by increasing the number concentration of cloud droplets, leading to a cooling effect on climate that is referred to as the Twomey effect^{1,2}. Here we show that anthropogenic aerosols exert an additional effect on cloud properties that is derived from changes in the spectral shape of the size distribution of cloud droplets in polluted air and acts to diminish this cooling. This finding could help to improve our understanding of the indirect aerosol effect and its treatment in climate modelling.

An equation commonly used for examination of the indirect aerosol effect is

$$r_e = \beta[3/(4\pi\rho)L/N]^{1/3} \quad (1)$$

where ρ is the water density, r_e is the effective radius, L is the cloud liquid-water content and N is the number concentration of cloud droplets. The parameter β is an increasing function of the relative dispersion (ε) of the cloud droplet size distribution (ratio of the standard deviation to the mean radius), which is well described^{3,4} by

$$\beta = (1 + 2\varepsilon^2)^{2/3}/(1 + \varepsilon^2)^{1/3} \quad (2)$$

An increase (decrease) in effective radius causes a decrease (increase) in cloud reflectivity^{1,2}. A prevailing assumption implicit in the evaluation of the indirect aerosol effect is that increasing aerosol loading does not alter ε or, hence, β . However, examination of data from field studies of the indirect aerosol effect shows that marine clouds classified as being polluted or having a continental origin generally have not only a larger N , but also a larger ε relative to unaffected marine clouds.

Figure 1 shows the dependence of ε and β on N . The points connected by lines represent cases identified by different investigators as evidence for the indirect effect. In each case, the points with lower N were characterized as background clouds and the higher points were characterized as similar clouds that had been perturbed by anthropogenic aerosols. Eleven of the 13 cases show an increase in ε that is concurrent with an increase in N , with negligible change to slight decreases in the other two; there is also a general increase in ε with N , as implied previously^{5–7}.

One explanation for the simultaneous increase in ε and N is that anthropogenic aerosols have a more complex chemical composition and a broader size distribution than marine aerosols, and that the more numerous small droplets formed in a

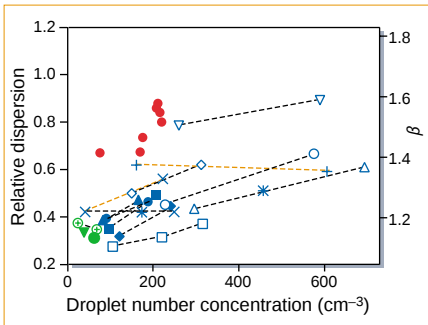


Figure 1 Relation between the relative dispersion of cloud droplet size distribution, ϵ , and the number concentration of cloud droplets, N . Symbols indicate programs and/or references from which the data points were derived. Connected points represent cases previously identified as evidence for an indirect aerosol effect. The parameter β is defined by equation (2). Green symbols (from ref. 8): triangle, FIRE, northeastern Pacific; crossed circles, SOCEX, Southern Ocean; filled circle, ACE1, Southern Ocean. Blue symbols: filled circles, ASTEX⁸, northeastern Atlantic; diamonds, SCMS⁹, Florida coast; filled triangles, Sounding⁹, ASTEX; filled squares, horizontal⁹, ASTEX; open inverted triangles, level 1; open upright triangles, level 2; open circles, level 3 — all from south-west of San Diego¹⁰; open diamonds, SCMS¹¹; stars, vertical, ASTEX¹²; plus signs, horizontal, ASTEX¹²; multiplication signs, ASTEX¹³; squares, INDOEX, Indian Ocean (G. M. McFarquhar, personal communication). Red circles, MAST^{6,14,15}, California coast.

polluted cloud compete for water vapour and broaden the droplet size distribution compared with clean clouds that have fewer droplets and less competition.

According to equations (1) and (2), an increase in ϵ acts to negate the effect of increased N on effective radius and cloud reflectivity. Because this effect has been largely neglected in estimates of the indirect aerosol effect, cooling by an indirect aerosol effect is likely to have been overestimated. From the data presented in Fig. 1, we estimate that a 15% increase in N at $N = 100 \text{ cm}^{-3}$ causes a total forcing that ranges between -0.19 and -0.93 W m^{-2} , which corresponds to a factor that is 10–80% lower than the -1.03 W m^{-2} calculated for the Twomey effect alone².

The effect of the enhancement in ϵ is evidently large enough to be considered in assessing the indirect aerosol effect, and understanding the relation between ϵ and N will help to reduce the large uncertainty inherent in this effect.

Yangang Liu, Peter H. Daum

Brookhaven National Laboratory, Upton,
New York 11973, USA
e-mail: lyg@bnl.gov

- Twomey, S. *Atmos. Environ.* **8**, 1251–1256 (1974).
- Charlson, R. J. *et al.* *Science* **255**, 423–430 (1992).
- Liu, Y. & Daum, P. H. *Geophys. Res. Lett.* **27**, 1903–1906 (2000).
- Liu, Y. & Daum, P. H. *Proc. 13th Int. Conf. On Clouds and Precipitation, Reno, USA* 586–589 (2000).
- Martin, G. M., Johnson, D. W. & Spice, A. J. *Atmos. Sci.* **51**, 1823–1842 (1994).
- Ackerman, A. S. *et al.* *J. Atmos. Sci.* **57**, 2684–2695 (2000).
- McFarquhar, G. M. & Heymsfield, A. J. *J. Geophys. Res.* **D 106**, 28675–28698 (2001).
- Yum, S. S. & Hudson, J. G. *Atmos. Res.* **57**, 81–104 (2001).
- Hudson, J. G. & Yum, S. S. *J. Atmos. Sci.* **54**, 2642–2654 (1997).

- Noonkester, V. R. *J. Atmos. Sci.* **41**, 829–845 (1984).
- Hudson, J. G. & Yum, S. S. *J. Atmos. Sci.* **58**, 915–926 (2001).
- Garrett, T. J. & Hobbs, P. V. *J. Atmos. Sci.* **52**, 2977–2984 (1995).
- Hudson, J. G. & Li, H. J. *Atmos. Sci.* **52**, 3031–3040 (1995).
- Noone, K. J. *et al.* *J. Atmos. Sci.* **57**, 2729–2747 (2000).
- Noone, K. J. *et al.* *J. Atmos. Sci.* **57**, 2748–2764 (2000).

Competing financial interests: declared none.

COMMUNICATIONS ARISING

Palaeoanthropology

Sahelanthropus or 'Sahelpithecus'?

Beginning with *Ramapithecus*, there has been a continued search for an ape-like hominid ancestor in the Miocene Epoch. *Sahelanthropus tchadensis* is an enigmatic new Miocene species, whose characteristics are a mix of those of apes and *Homo erectus* and which has been proclaimed by Brunet *et al.* to be the earliest hominid¹. However, we believe that features of the dentition, face and cranial base that are said to define unique links between this Toumaï specimen and the hominid clade are either not diagnostic or are consequences of biomechanical adaptations. To represent a valid clade, hominids must share unique defining features², and *Sahelanthropus* does not appear to have been an obligate biped.

We consider the following features that are proposed by Brunet *et al.* to constitute links with the hominid clade. First, they note that the canine is small. The canine breadth (the length is not given) is similar to the chimpanzee mean, being within the range of both chimpanzee and gorilla females and of chimpanzee males. The crown is low and the root is narrow relative to the crown, suggesting that Toumaï might have been female (canine area is a more reliable sex indicator than brow ridges); however, the postcanine teeth are all large compared with chimpanzees — as in several Miocene ape females.

The canine shows apical wear and has a thin strip of wear along the distal edge of the crown, which reaches the crown base. Although Brunet *et al.* conclude that the tooth was not used in honing, we find this difficult to reconcile with the details that they provide. Decades ago, when Miocene primate jaws with small canines and enlarged postcanine teeth were found, they were given distinct names (*Kenyapithecus* and *Ramapithecus*, for example) and were described as the earliest hominids because (it was assumed) the canine honing function had, by then, been replaced by tools (it was also assumed that they were bipeds)³. These specimens turned out to be female apes.

Second, the specimen has a large, continuous supraorbital torus, and the authors

claim that there are other facial similarities to *Homo*. However, the facial similarities are mostly not with early hominids but with Pleistocene *Homo*, and therefore do not provide any phylogenetic information (no evidence hints that *Homo erectus* could be 6–7 million years old). There is little subnasal prognathism because the canines are small and the subnasal region is short, and the closely packed anterior dentition, crowded together because of the expanded postcanine teeth, explains the absence of diastemata. The vertical height of the impressive supraorbitals is greater than in any extant ape or australopithecine, and can only be matched in *Homo erectus* and in a few later humans.

The supraorbital size is attributed to strong sexual selection¹, which we consider unlikely. The size and form of the supraorbital structures are probably a mechanical response⁴ to strain from anterior tooth loading in the region above the orbits, concentrated by the flexed frontofacial angle. The biomechanical model of the supraorbital region⁵ would predict that an orthognathic face such as that of *Sahelanthropus*, combined with a low forehead, creates the potential for greater strain during anterior tooth loading than would a prognathic face with a higher forehead, as in African apes. Significant force during anterior tooth use is indicated by the expanded posterior temporalis musculature — this muscle forms a sagittal crest that meets the nuchal crest very high on the posterior vault, in a gorilla-like morphology that far exceeds the much weaker-muscled chimpanzee condition. The supraorbital torus is the bony response to strain.

Third, Brunet *et al.* infer that the intermediate thickness of the specimen's postcanine enamel is also an important link. However, thickened enamel relative to chimpanzees would be expected whatever the phylogenetic relations of *Sahelanthropus*, not only because of the other adaptations to a diet that requires powerful mastication, but also because thickened enamel is a plesiomorphic condition.

Fourth, the authors assign an anterior position to the foramen magnum on the basis of its front edge meeting the bicarotid and biporion chords, claiming that in chimpanzees the foramen magnum is posterior to these chords and that this positioning reflects their posture. STS 5, an obligate biped, does not differ from some chimpanzees by these criteria, however. Neither does this apply to the biporion chord of some chimpanzees, according to the position of the foramen magnum determined from photographs of 70 adult chimpanzee cranial bases aligned with the Frankfurt horizontal (J. Ahern, personal communication). This alignment is important in the determination, but we have no information

on how the *Sahelanthropus* position was assessed. Moreover, the anterior edge of the foramen is far from the back of the *Sahelanthropus* third molar, in contrast to hominids and similar to chimpanzees and female gorillas.

There are many other features that link the specimen with chimpanzees, gorillas or both, to the exclusion of hominids. Most significantly, the nuchal plane is long, flat and angled at about 55° to the Frankfurt horizontal: “relatively longer than in *Pan* [and] *Gorilla* ... and with crests as marked as those of *Gorilla*”¹. This describes the posterior cranial vault of a small quadrupedal ape with a powerful masticatory complex.

Because the face is orthognathic rather than prognathic and the anterior teeth are small, posture is the only credible explanation of this nuchal anatomy. It is evident that *Sahelanthropus* did not habitually hold its head in an upright position over the spine and was not an obligate biped. This contrast with all known hominids is itself sufficient to exclude *Sahelanthropus* from the hominid clade as we currently understand it.

We believe that *Sahelanthropus* was an ape living in an environment that was later inhabited by australopithecines and, like them, it adapted with a powerful masticatory complex. A penecontemporary primate with a perfect and well-developed postcranial adaptation to obligate bipedalism⁶ is more likely to have been an early hominid.

Milford H. Wolpoff*, Brigitte Senut†, Martin Pickford‡, John Hawks§

*Paleoanthropology Laboratory, Department of Anthropology, University of Michigan, Ann Arbor, Michigan 48109-1382, USA
e-mail: wolpoff@umich.edu

†Laboratoire de Paléontologie du Muséum National d'Histoire Naturelle, and ‡Chaire de Paléanthropologie et de Préhistoire du Collège de France, UMR 8569 CNRS, 75005 Paris, France
§Department of Anthropology, University of Wisconsin, Madison, Wisconsin 53706, USA

1. Brunet, M. *et al.* *Nature* **418**, 145–151 (2002).
2. Contra Wood, B. *Nature* **418**, 133–135 (2002).
3. Pilbeam, D. R. *Am. J. Phys. Anthropol.* **25**, 1–6 (1966).
4. Endo, B. *Jpn. Facul. Sci., Univ. Tokyo, Sec. 5*, **3**, 1–106 (1966).
5. Russell, M. D. *Curr. Anthropol.* **26**, 337–360 (1985).
6. Senut, B. *et al.* *C. R. Acad. Sci. Paris* **332**, 137–144 (2001).

Brunet et al. reply — In 1925, when Dart described *Australopithecus africanus*¹ as a hominid, critics interpreted it as a juvenile gorilla^{2–4}. Last year, Wolpoff's colleagues (B.S. and M.P.) claimed that their Kenyan fossil *Ororin* was a direct ancestor of *Homo*⁵, and now Wolpoff *et al.* conclude that *Sahelanthropus* was an ape (specifically, a female gorilla ancestor⁶) — a belief that, to our knowledge, is not supported by published or unpublished data.

Overlooking their flippant taxonomic proposal (the genus name ‘*Sahelpithecus*’), which disregards the requirement for a new genus to have a type species and description⁷, we disagree with their (presumably more serious) opinions on the morphology and phylogeny of the Toumaï fossil.

Because the Toumaï fossil is the earliest known hominid ancestor⁸, it is not surprising that it bears primitive characters. Following modern systematic practice, we used newly evolved characters (rather than shared primitive characters) to establish phylogenetic relationships⁹. Those who ignore these derived characters and concentrate on primitive ones will reach the conclusion that early hominids, including *Ororin*, are related to modern apes. This has not been in dispute since Huxley and Darwin. For Wolpoff *et al.* to revert to the use of primitive characters in an attempt to undermine a clear statement of affinity of Toumaï is curious.

Wolpoff *et al.* make several erroneous assertions about the cranial face and base. For example, they mischaracterize the configuration of the face in *S. tchadensis*, claiming that supraorbital size is directly related to postcanine tooth size and/or to masticatory forces. However, experimental and developmental investigations^{9,10} have shown that strains caused by mastication in the brow ridge of orthognathic and prognathic primates are always tiny, much too small to engender bone-growth responses to loading. Instead, large brow ridges grow because of facial projection relative to the cranial base¹¹.

Wolpoff *et al.* also obfuscate the facial similarities to *Homo*. We did not suggest that *Homo erectus* is 6–7 million years old — the point with *Homo* was comparative, rather than phylogenetic. Relying on measurements of our published photographs of the distorted original, Wolpoff *et al.* wrongly assert that the nuchal plane is angled at about 55° to the Frankfurt horizontal. Undistorted, the nuchal plane's angulation is outside the range of chimpanzees and within the range of fossil hominids¹². This configuration is nothing like that of any quadrupedal ape, with or without a powerful masticatory complex (which *Sahelanthropus* lacks, contrary to the assertions of Wolpoff *et al.*).

These authors not only misrepresent the specimen's morphology, but also fail to identify a single character to support their suggestion that Toumaï is a gorilla rather

than a hominid ancestor. They interpret our description of distal dentin exposure of the upper canine as evidence of honing wear (roughly equivalent to describing an African millet pestle as a Samurai sword). The Toumaï canine is not honing because it does not display the sharpened distal edge that is shared by all apes. Rather, this tooth is similar to those of later hominids in both size and proportion to the post-canine teeth.

In a modern example of how to miss the morphology between measuring points, Wolpoff *et al.* argue that the size of the Toumaï canine is ape-like. It is well known that early hominid and modern ape canine buccolingual diameters overlap in size. But, as Broom and Robinson¹³ noted in their assessment of Zuckerman's failed attempt to sideline *Australopithecus* 50 years ago: “If ... the affinities of an animal are to be determined by the size and indices of its teeth, and not by their structure, a horse may have to be put in the same group as a cow.” In its relative size, morphology and wear, the Toumaï canine is derived in the hominid direction relative to any ape.

This phylogenetic signal is significant. Ignoring it in favour of a belief based on *Ororin* and primitive characters is unjustified, particularly as the phylogenetic position of the *Ororin* fossils remains uncertain. Wolpoff *et al.* have described no derived ape feature of *S. tchadensis*, nor have they disproved any derived features that this species shares with later hominids⁸. Any alternative phylogenetic hypothesis should be based on explicit, supporting derived characters of Toumaï.

Michel Brunet (for the authors)

Faculté des Sciences et CNRS UMR 6046, Université de Poitiers, 86022 Poitiers, France
e-mail: michel.brunet@univ-poitiers.fr

1. Dart, R. A. *Nature* **115**, 195–199 (1925).
2. Keith, A. *Nature* **115**, 234 (1925).
3. Duckworth, W. L. H. *Nature* **115**, 236–237 (1925).
4. Keith, A. *Nature* **159**, 377 (1947).
5. Senut, B. *et al.* *C. R. Acad. Sci. Paris* **332**, 137–144 (2001).
6. Fléaux, R. *Sci. Avenir Paris* **666**, 12 (2002).
7. International Commission on Zoological Nomenclature *International Code of Zoological Nomenclature* 4th edn (Int. Trust Zool. Nomenclature, London, 1999).
8. Brunet, M. *et al.* *Nature* **418**, 145–151 (2002).
9. Hylander, W. L. & Johnson, K. R. in *The Biological Mechanisms of Tooth Movement* (ed. Davidovitch, Z.) 559–569 (Ohio State Univ. Coll. Dentistry, Columbus, 1992).
10. Ross, C. R. *Am. J. Phys. Anthropol.* **116**, 108–139 (2001).
11. Lieberman, D. E. in *Development, Growth and Evolution: Implications for the Study of the Hominid Skeleton* (eds O'Higgins, P. & Cohn, M.) 86–122 (Academic, London, 2000).
12. Kimbel, W. H., White, T. D. & Johanson, D. C. *Am. J. Phys. Anthropol.* **64**, 337–388 (1984).
13. Broom, R. & Robinson, J. T. *Mem. Transvaal Mus.* **6** (1952).

brief communications is intended to provide a forum for both brief, topical reports of general scientific interest and technical discussion of recently published material of particular interest to non-specialist readers. Priority will be given to contributions that have fewer than 500 words, 10 references and only one figure. Detailed guidelines are available on *Nature's* website (www.nature.com) or on request from nature@nature.com

Somatic cell nuclear transfer

I. Wilmut*, N. Beaujean*, P. A. de Sousa*, A. Dinnyes*†, T. J. King*, L. A. Paterson*, D. N. Wells‡ & L. E. Young*†

* Roslin Institute, Roslin, Midlothian EH25 9PS, UK

† AgResearch Ltd, PB3123, Hamilton, New Zealand

Cloning by nuclear transfer from adult somatic cells is a remarkable demonstration of developmental plasticity. When a nucleus is placed in oocyte cytoplasm, the changes in chromatin structure that govern differentiation can be reversed, and the nucleus can be made to control development to term.

Cloning by present methods is very inefficient owing to the extraordinary demands placed on the oocyte cytoplasm in reprogramming a somatic nucleus rather than a sperm nucleus. The cumulative loss observed throughout development is assumed to reflect inappropriate expression of many genes whose harmful effect is exerted at different stages of development. These fundamental limitations to cloning are being addressed by analyses of the underlying cellular mechanisms. In time, this information may be used in the development of treatments to cause cells of one phenotype to 'transdifferentiate' to another.

Several laboratories have used a variety of somatic cell types to create cloned sheep, cattle, mice, pigs, goats, rabbits and cats (results are summarized in ref. 1). However, those laboratories have failed so far to obtain offspring in other species, including the rat, rhesus monkey and dog. Consistent among all published research is that only a small proportion of embryos reconstructed using adult or fetal somatic cells developed to become live young, typically between 0 and 4%. The low overall success rate is the cumulative result of inefficiencies at each stage of the process (Fig. 1).

Fetal loss and pathologies of cloned offspring

In addition to embryonic loss, somatic cell nuclear transfer is also associated with very high rates of fetal, perinatal and neonatal loss, and production of abnormal offspring. Not all of these effects are due solely to nuclear transfer as similar problems are reported after embryo culture².

Typically, at least one-third of the cattle and sheep confirmed pregnant with cloned embryos lose their fetuses during gestation¹. Abnormal development of the placenta, including vascular reduction, is a principal contributor to loss particularly during early pregnancy in sheep and cattle³. It may also contribute to some of the defects reported in neonates⁴. In cattle the rate of loss is also increased in the second and third trimesters of pregnancy after nuclear transfer (compared with *in vitro* fertilization (IVF)), with greater losses when adult rather than fetal or embryonic nuclei are used⁵. The overaccumulation of placental fluid in hydroallantois occurs rarely in natural cattle pregnancies, but can affect up to 2% and 40% of pregnancies established with IVF⁶ and cloned embryos⁷, respectively. In cloned mice, the placentae are often 2–3-fold heavier than from natural mating⁸, although a lack of vascularization has not been reported.

Many cloned offspring die within the first 24 h of birth (Table 1). Common anomalies include respiratory distress, increased birth weight and major cardiovascular abnormalities that can result in gross distension of the liver and dilated major vessels. Some of these anomalies may be secondary effects of another abnormality. The production of oversized offspring after nuclear transfer in ruminants mimics the findings following other embryo manipulations

when increased birth weight is the most readily recognized characteristic of the 'large offspring syndrome'². Other common characteristics include prolonged gestation, fluid accumulation, enlargement of organs, sluggish onset of labour and difficulty in breathing.

There has not been an adequate assessment of post-natal development of clones. Abnormalities that have been described include failure of the immune system, structural abnormalities of the brain, digestive dysfunction, enteritis and umbilical infections. There is a suggestion that the fate of cloned individuals may be influenced by genetic background or donor cell type. In two independent studies in which the strains of mice used differed, mice cloned from cumulus cells became obese in adult life⁹ whereas those from Sertoli cells of immature mice died at an unusually early age¹⁰. The fact that the obesity was not heritable after sexual reproduction provides direct evidence that failures associated with cloning are epigenetic⁹. By contrast, physiological studies of cloned calves suggest normality, at least for the tests examined¹¹. However, the eldest of these animals was only 4 years at the time of the tests and it is important that studies extend throughout a lifespan, which is potentially more than 15 years.

Cellular factors influencing outcome

Maintenance of ploidy

Normal development of cloned embryos depends on maintaining normal ploidy by coordinating the cell cycles of the recipient oocyte and donor nucleus (see ref. 12 for review). Three effective strategies have been defined. When using mammalian oocytes arrested at metaphase II as the recipient cells, normal ploidy is maintained if the donor nucleus is awaiting DNA replication (G0 or G1). Alternatively, advantage may be taken of the fact that ploidy is also changed if chromosomes are shed in a polar body. Following transfer of G2/M-phase nuclei (4C) into enucleated metaphase II mouse oocytes¹³, expulsion of a polar body returns the reconstructed embryo to normal ploidy of 2C, provided that cytochalasin B is omitted from the culture medium. In the third strategy, an

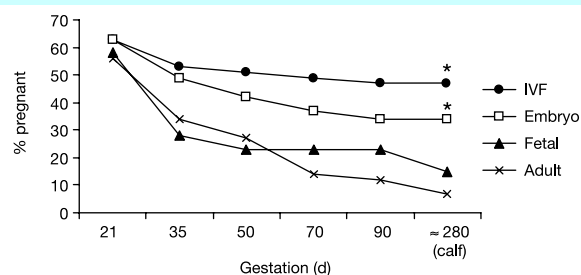


Figure 1 A comparison of survival to term of bovine embryos produced either by *in vitro* fertilization or nuclear transfer from embryos, fetal or adult fibroblasts. (Data from ref. 5.) *Including one twin calving from a single embryo transfer.

† Present addresses: Research Group for Animal Breeding, Hungarian Academy of Sciences & Szent István University, Hungary (A.D.); and School of Human Development, Division of Reproductive Medicine, Queens Medical Centre, Nottingham NG7 2UH, UK (L.E.Y.).

interphasic donor nucleus may be at any stage of the cycle if the metaphase II oocyte is activated before transfer and M-phase promoting factor (MPF) levels decline. However, cytoplasts with basal levels of MPF may not be optimal for reprogramming somatic nuclei¹⁴. Although these general principles were established quickly, it seems probable that there is much to learn of the detailed mechanisms in different species. Rabbits were successfully cloned only when adjustments were made to accommodate for the rapid progression through the first cell cycle in that species¹⁵.

Donor cell cycle stage

Although few precise comparisons have been made, there appears to be a 'window of opportunity' in the cycle of donor cells, spanning G2/M/G1/G0 phases, when nuclear reprogramming is more effective. Dolly (the first successful clone of an adult animal) was created from a donor cell induced to quiescence by serum deprivation¹⁶, and since then many other offspring have been produced using cells that were expected to be quiescent. Three studies have now carefully selected cells in G1 and contrasted them with cells in G0 (refs 7, 17, 18). All three studies obtained calves from cells in G1, although in one case no calves were derived with fetal cells using the well-proven techniques in G0 (ref. 17). The other two studies, however, showed a 3–7-fold increase in the production of viable offspring with embryos derived from adult quiescent cells (G0)^{7,18}. Donor cells at metaphase may be considered optimally compatible with metaphase II oocytes and have resulted in improved rates of blastocyst development using murine embryonic stem (ES) cells, as opposed to interphasic nuclei; however, the post-implantation development was then characterized by a high mortality¹⁹. Determining optimal cell cycle stage of the donor cell is of great practical importance and would also focus mechanistic research.

Delayed activation

After nuclear transfer, an activation stimulus triggers development. It was initially considered that the direct exposure of chromosomes to factors in the oocyte cytoplasm before activation²⁰ and allowing time for the re-assembly of the nuclear apparatus required for the cell cycle²¹ might aid development. Although delaying oocyte activation after nuclear transfer can enhance development in cattle and mice, it is not essential in these species^{14,22}.

Stage of development of donor cells

In general there is a decrease in the proportion of embryos becoming offspring when donor nuclei are obtained from later stages of development. The change is assumed to reflect the degree of cellular differentiation and epigenetic constraints. The proportion of mouse embryos that developed to offspring with transferred blastomere nuclei from 2-, 4- and 8-cell stage embryos decreased from 22%, 14% and 8%, respectively²³. Embryonic survival decreases further with pluripotent mouse ES cells (2–11%) and somatic cells in G0/G1 (1–3% (ref. 1)). The reduction in efficiency occurs later in cattle, as the proportion of transferred bovine cloned embryos that developed into live offspring from

morulae, fetal and adult fibroblasts was 28%, 13% and 5%, respectively (Fig. 1). Differences between species are associated with the later onset of major transcription from the embryonic genome in ruminants than mice²⁴. This may mean that developmental changes in chromatin structure take place later in ruminant development and so less remodelling is required in cleavage stage embryos. Additionally, after nuclear transfer more time and cell cycles are available for remodelling before transcription begins in ruminants. Recent observations in mice using fetal neural cells as nuclear donors have also suggested a loss of developmental totipotency with neural differentiation²⁵. However, terminal differentiation-associated DNA rearrangements²⁵ may not be incompatible with nuclear totipotency as evidenced by successful cloning from terminally differentiated lymphocytes²⁶. In mice, differences in the pattern of fetal loss when using hybrid and inbred donor cells were particularly apparent with ES cells as nuclear donors, and efficiency was also lower when inbred cumulus cells were used²⁷.

Molecular mechanisms

The unique pattern of mortality during embryonic, fetal and neonatal stages in clones apparently reflects the inappropriate expression of genes whose effect is exerted at different stages of development and can be fatal. Perturbations in the expression of imprinted genes associated with phenotypic defects have now been observed in sheep, cattle and mice. Many imprinted genes have critical roles in regulating growth during fetal development; furthermore, the common characteristics of the large offspring syndrome were recognized as being similar to those clinical and experimental situations in which the expression or effect of H19 and/or insulin-like growth factor II (IGF2) is perturbed². Expression of IGF2 receptor (IGF2R) was reduced in the liver, kidney, heart and muscle of unusually large sheep fetuses resulting from embryo culture²⁸. The change in expression was associated with a loss of methylation at a differentially methylated region involved in regulation of gene imprinting. Unexpectedly, there was no change in expression of IGF2 or H19.

Mouse fetuses derived by nuclear transfer from ES cells display aberrant expression of *H19*, *Igf2* and other imprinted genes known to have important roles in fetal or placental development. Some mice cloned from ES cells survive until adulthood and appear phenotypically normal despite epigenetic aberrations in at least one of the six imprinted genes that were examined²⁹. Evidently 'normal' life is possible despite considerable variation in gene expression, and this probably depends on the particular gene affected, its degree and nature of perturbation, effects of redundancy, and on species-specific variation in susceptibility to any single imprinted gene disruption. However, rather than being the consequence of nuclear transfer, the imprinting status of the cloned mice examined in ref. 29 may be due to the epigenetic instability inherent to ES cell lines. When somatic cells were used as nuclear donors in mice, imprinting errors were not detected in the phenotypically normal fetus; however, the level of expression of several

Table 1 Summary of the pathologies described in cloned fetuses and offspring

Species	Type of embryo manipulation	Embryo loss	Gestation length	Placental abnormalities	Fetal size	Respiratory/cardiovascular dysfunction	Organ dysplasia	Perinatal mortality	Post-natal development
Cow	Nuclear transfer	High	Prolonged	Common	Increases	Common	Common	Raised	Altered
	Other	High	Prolonged	Common	Increases	Occasional	Occasional	Raised	Altered
Sheep	Nuclear transfer	High	Prolonged	Common	Increases	Common	Common	Raised	Altered
	Other	High	Prolonged	Common	Increases	-	Occasional	Raised	Altered
Goat	Nuclear transfer	High	Prolonged	None observed	Normal	No	No	Normal	Normal
	Other	High	Normal	None observed	Normal	No	No	Normal	Normal
Pig	Nuclear transfer	High	Prolonged	None observed	Reductions	Occasional	Occasional	Raised	Altered
	Other	High	Normal	None observed	Reductions	No	No	Raised	Normal
Mouse	Nuclear transfer	High	Caesareans	Common	Altered	Common	-	Raised	Altered
	Other	High	-	-	Reduced	-	-	-	Altered

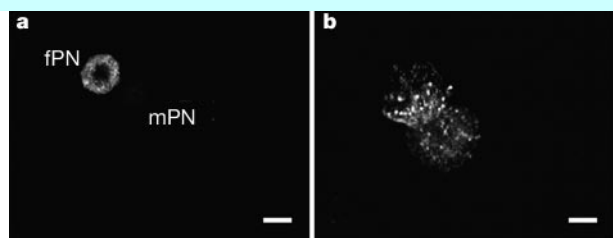


Figure 2 Methyl cytosine immunostaining of pronuclear DNA in mouse and sheep zygotes. **a**, Mouse 1-cell embryo 14 h after fertilization. Only the female pronucleus (fPN) is stained. mPN indicates position of the male pronucleus. **b**, Sheep 1-cell embryo 22 h after fertilization. In marked contrast to the mouse, the paternal pronucleus does not undergo demethylation and both pronuclei are stained. Scale bar, 10 μ m.

genes was elevated and imprinting errors were observed in their overgrown placentae³⁰.

Apart from the improper expression of imprinted genes, the dysregulation of non-imprinted genes also occurs in clones³¹. However, two measures of the correct reprogramming of gene expression have been reported elsewhere. In female somatic cells, reactivation of the inactive X chromosome can occur in early embryos after nuclear transfer, and reprogramming of telomerase activity can allow restoration of telomere length in cloned animals (reviewed in ref. 32).

Epigenetic regulation of gene expression

Errors leading to inappropriate expression could arise at any of the levels at which regulation of gene expression occurs, including the organization of the nucleus, chromatin structure and the availability of regulatory molecules. Chromatin proteins regulate access of the transcription machinery to chromosomal DNA and so ultimately determine the pattern of gene expression. This chromosomal infrastructure is far more dynamic than had been understood previously, adding to the impression that nuclear and chromosome structure are in a continual state of flux³³. A large number of structural proteins and enzymes that modify chromatin structure have been identified and are principal candidates for regulating early reprogramming events.

The linker histone H1 may be involved in the regulation of gene expression in early embryos (reviewed in ref. 34). The somatic form of histone H1 is apparently absent or in very low concentration in oocytes and early embryos, but becomes detectable when major transcriptional activation of the embryonic genome occurs. Somatic H1 is lost from most mouse nuclei soon after transfer, but the cell cycle stage of donor and recipient cells influences the rate of loss³⁵. Enzymatic modifications of histones include phosphorylation, methylation, acetylation and ubiquitination, or the removal of these modifications³⁶. These modifications are recognized by other structural proteins and enzymes, which together may stabilize the pattern of gene expression.

Chromatin remodelling is also achieved by multisubunit ATP-dependent enzymes (reviewed in ref. 36). The helicase activity of these enzymes is able to expose DNA or redistribute nucleosomes in a tissue-specific manner. The loss of a principal component of the basal transcriptional complex from somatic nuclei incubated in frog oocyte extract provided the first indication that members of the SWI/SNF family of enzymes may have roles in the development of cloned embryos³⁷.

Finally, methylation of CpG dinucleotides in chromosomal DNA provides a genome-wide means of regulation, usually associated with the inheritance of lineage-specific gene silencing between cell generations (reviewed in ref. 38). Widespread disruptions in DNA methylation have now been described in cloned embryos of mice and cattle (reviewed in ref. 39), suggesting that present methods of

cloning are disruptive and that the effects are complex. In cloned cattle embryos, the highly methylated status of the donor cells was retained through several cell divisions. There was great variation between cloned embryos consistent with the variation in stage and cause of death of cloned embryos. In contrast, in the pig, changes in methylation of short interspersed element sequences were similar to those observed after *in vitro* fertilization, although at the 4–8-cell stage both were more extensively methylated than in embryos fertilized *in vivo* (reviewed in ref. 39).

Together these observations emphasize the need for further studies of different regions of the genome with different donor cells and methods of nuclear transfer in several species. It is critical not to assume that mammalian pre-implantation events (and therefore reprogramming of a somatic nucleus) are the same in all species. The rapid demethylation of the paternal genome that is seen in several species does not occur in the sheep (Fig. 2: N.B., I.W., R. R. Meehan and L.E.Y., unpublished data).

The future

As the fate of cloned embryos is determined by molecular events within hours of nuclear transfer, it is disappointing that so little is known about these events during the early development of cloned embryos. During somatic cell nuclear transfer a great deal is asked of the molecular mechanisms that have evolved to regulate fertilization and pregnancy. Viewed in this light, it is still surprising that somatic cell cloning ever produces viable offspring. Although some improvements in efficiency are to be expected from optimization of present procedures, greater benefits might be expected from intervention to assist reprogramming of the transferred nucleus. At present the means to enhance the success of nuclear transfer are not known, but may involve the use of remodelling complexes and factors that remove somatic epigenetic modifications before transfer. In addition to application of this information in nuclear transfer, new understanding of mechanisms that regulate developmental plasticity will lead to methods to change cells of one phenotype to another as a means of providing histocompatible cells for treatment of degenerative diseases. A new era of developmental biology and regenerative medicine awaits. □

doi:10.1038/nature01079.

- Wilmot, I. & Peterson, L. A. Somatic cell nuclear transfer (cloning) efficiency [online] (<http://www.roslin.ac.uk/public/webtablesGR.pdf>) (2002).
- Young, L. E. & Fairburn, H. R. Improving the safety of embryo technologies: possible role of genomic imprinting. *Theriogenology* **53**, 627–648 (2000).
- Hill, J. R. *et al.* Evidence for placental abnormality as the major cause of mortality in first-trimester somatic cell cloned bovine fetuses. *Biol. Reprod.* **63**, 1787–1794 (2000).
- Hill, J. R. *et al.* Clinical and pathologic features of cloned transgenic calves and fetuses (13 case studies). *Theriogenology* **51**, 1451–1465 (1999).
- Heyman, Y. *et al.* Frequency and occurrence of late-gestation losses from cattle cloned embryos. *Biol. Reprod.* **66**, 6–13 (2002).
- Hasler, J. F. *et al.* Production, freezing and transfer of bovine IVF embryos and subsequent calving results. *Theriogenology* **43**, 141–152 (1995).
- Wells, D. *et al.* Proc. 32nd Ann. Conf. Soc. Reprod. Biol. Abstract No. 25 (2001).
- Tanaka, S. *et al.* Placentomegaly in cloned mouse concepti caused by expansion of the spongiotrophoblast layer. *Biol. Reprod.* **65**, 1813–1821 (2001).
- Tamashiro, K. L. *et al.* Cloned mice have an obese phenotype not transmitted to their offspring. *Nature Med.* **8**, 262–267 (2002).
- Ogonuki, N. *et al.* Early death of mice cloned from somatic cells. *Nature Genet.* **30**, 253–254 (2002).
- Lanza, R. P. *et al.* Cloned cattle can be healthy and normal. *Science* **294**, 1893–1894 (2001).
- Campbell, K. H., Loi, P., Otaegui, P. J. & Wilmot, I. Cell cycle co-ordination in embryo cloning by nuclear transfer. *Rev. Reprod.* **1**, 40–46 (1996).
- Ono, Y., Shimozawa, N., Ito, M. & Kono, T. Cloned mice from fetal fibroblast cells arrested at metaphase by a serial nuclear transfer. *Biol. Reprod.* **64**, 44–50 (2001).
- Wakayama, T. & Yanagimachi, R. Effect of cytokinesis inhibitors, DMSO and the timing of oocyte activation on mouse cloning using cumulus cell nuclei. *Reproduction* **122**, 49–60 (2001).
- Chesne, P. *et al.* Cloned rabbits produced by nuclear transfer from adult somatic cells. *Nature Biotechnol.* **20**, 366–369 (2002).
- Wilmot, I., Schnieke, A. E., McWhir, J., Kind, A. J. & Campbell, K. H. S. Viable offspring derived from fetal and adult mammalian cells. *Nature* **385**, 810–813 (1997).
- Kasinathan, P., Knott, J. G., Wang, Z., Jerry, D. J. & Robl, J. M. Production of calves from G1 fibroblasts. *Nature Biotechnol.* **19**, 1176–1178 (2001).
- Wells, D. *et al.* Proc. Curr. Status Perspec. Cloning Stud. 73 (National Institute of Agrobiological Sciences, Tsukuba, Japan, 2001).

19. Zhou, Q., Jouneau, A., Brochard, V., Adenot, P. & Renard, J. P. Developmental potential of mouse embryos reconstructed from metaphase embryonic stem cell nuclei. *Biol. Reprod.* **65**, 412–419 (2001).
20. Campbell, K. H. S., McWhir, J., Ritchie, W. A. & Wilmut, I. Sheep cloned by nuclear transfer from a cultured cell line. *Nature* **380**, 64–66 (1996).
21. Wakayama, T., Perry, A. C. F., Zuccotti, M., Johnson, K. R. & Yanagimachi, R. Full-term development of mice from enucleated oocytes injected with cumulus cell nuclei. *Nature* **394**, 369–374 (1998).
22. Kubota, C. *et al.* Six cloned calves produced from adult fibroblast cells after long-term culture. *Proc. Natl Acad. Sci. USA* **97**, 990–995 (2000).
23. Cheong, H. T., Takahashi, Y. & Kanagawa, H. Birth of mice after transplantation of early cell-cycle-stage embryonic nuclei into enucleated oocytes. *Biol. Reprod.* **48**, 958–963 (1993).
24. Thompson, E. M. Chromatin structure and gene expression in the preimplantation mammalian embryo. *Reprod. Nutr. Dev.* **36**, 619–635 (1996).
25. Yamazaki, Y. *et al.* Assessment of the developmental totipotency of neural cells in the cerebral cortex of mouse embryo by nuclear transfer. *Proc. Natl Acad. Sci. USA* **98**, 14022–14026 (2001).
26. Hochedlinger, K. & Jaenisch, R. Monoclonal mice generated by nuclear transfer from mature B and T cells. *Nature* **415**, 967–969 (2002).
27. Wakayama, T. & Yanagimachi, R. Mouse cloning with nucleus donor cells of different age and type. *Mol. Reprod. Dev.* **58**, 376–383 (2001).
28. Young, L. E. *et al.* Epigenetic change in IGF2R is associated with fetal overgrowth after sheep embryo culture. *Nature Genet.* **27**, 153–154 (2001).
29. Humpherys, D. *et al.* Epigenetic instability in ES cells and cloned mice. *Science* **293**, 95–97 (2001).
30. Inoue, K. *et al.* Faithful expression of imprinted genes in cloned mice. *Science* **295**, 297 (2002).
31. Daniels, R., Hall, V. & Trounson, A. O. Analysis of gene transcription in bovine nuclear transfer embryos reconstructed with granulosa cell nuclei. *Biol. Reprod.* **63**, 1034–1040 (2000).
32. Rideout, W. M., Eggan, K. & Jaenisch, R. Nuclear cloning and epigenetic reprogramming of the genome. *Science* **293**, 1093–1098 (2001).
33. Misteli, T. Protein dynamics: implications for nuclear architecture and gene expression. *Science* **291**, 843–847 (2001).
34. Clarke, H. J., McLay, D. W. & Mohamed, O. A. Linker histone transitions during mammalian oogenesis and embryogenesis. *Dev. Genet.* **22**, 17–30 (1998).
35. Bordignon, V., Clarke, H. J. & Smith, L. C. Factors controlling the loss of immunoreactive somatic histone H1 from blastomere nuclei in oocyte cytoplasm: a potential marker of nuclear reprogramming. *Dev. Biol.* **233**, 192–203 (2001).
36. Nakao, M. Epigenetics: interaction of DNA methylation and chromatin. *Gene* **278**, 25–31 (2001).
37. Kikyo, N., Wade, P. A., Guschin, D., Ge, H. & Wolffe, A. P. Active remodeling of somatic nuclei in egg cytoplasm by the nucleosomal ATPase ISWI. *Science* **289**, 2360–2362 (2000).
38. Robertson, K. D. & Wolffe, A. P. DNA methylation in health and disease. *Nature Rev. Genet.* **1**, 11–19 (2000).
39. Fairburn, H. R., Young, L. E. & Hendrich, B. D. Epigenetic reprogramming: how now, cloned cow. *Curr. Biol.* **12**, R68–R70 (2002).

Acknowledgements We each acknowledge discussions with colleagues during the development of these ideas and are grateful for funding from the governments, charities and commercial agencies acknowledged in our research papers.

Correspondence and requests for materials should be addressed to I.W. (e-mail: ian.wilmut@bbsrc.ac.uk).

Crystal structure of bacterial multidrug efflux transporter AcrB

Satoshi Murakami^{*†§}, Ryosuke Nakashima^{*}, Eiki Yamashita[‡] & Akihito Yamaguchi^{*†§}

^{*} Department of Cell Membrane Biology, Institute of Scientific and Industrial Research, Osaka University, Ibaraki, Osaka 567-0047, Japan

[†] Faculty of Pharmaceutical Science, 1-6 Yamada-oka, Osaka University, Suita, Osaka 565-0871, Japan

[‡] Institute for Protein Research, Osaka University, 3-2 Yamada-oka, Suita 565-0871, Japan

[§] CREST, Japan Science and Technology Corporation, Osaka 567-0047, Japan

AcrB is a major multidrug exporter in *Escherichia coli*. It cooperates with a membrane fusion protein, AcrA, and an outer membrane channel, TolC. We have determined the crystal structure of AcrB at 3.5 Å resolution. Three AcrB protomers are organized as a homotrimer in the shape of a jellyfish. Each protomer is composed of a transmembrane region 50 Å thick and a 70 Å protruding headpiece. The top of the headpiece opens like a funnel, where TolC might directly dock into AcrB. A pore formed by three α -helices connects the funnel with a central cavity located at the bottom of the headpiece. The cavity has three vestibules at the side of the headpiece which lead into the periplasm. In the transmembrane region, each protomer has twelve transmembrane α -helices. The structure implies that substrates translocated from the cell interior through the transmembrane region and from the periplasm through the vestibules are collected in the central cavity and then actively transported through the pore into the TolC tunnel.

The emergence of bacterial multidrug resistance is an increasing problem in the treatment of infectious diseases. Multidrug resistance often results from the overexpression of a multidrug efflux system^{1,2}. Genome sequence analysis of microorganisms has revealed the presence of several putative drug efflux transporter genes³. Among them, the AcrAB–TolC system is constitutively expressed in *E. coli* and is largely responsible for the intrinsic resistance of *E. coli* against dyes, detergents and most lipophilic antibiotics⁴. The AcrAB–TolC system is composed of a resistance-nodulation-cell division (RND)⁵ transporter, AcrB, a membrane fusion protein (MFP)⁶, AcrA, and a multifunctional outer membrane channel, TolC⁷; it transports a wide variety of noxious compounds from the cell directly into the medium, bypassing the periplasm^{8,9}. The substrate specificity of the AcrAB system is

unusually broad. It extrudes cationic, neutral and anionic substances¹⁰.

In addition, the AcrAB system pumps out some β -lactams with multiple charged groups, which were experimentally shown not to traverse the cytoplasmic membrane. These substrates are thus likely to be captured from the periplasm¹⁰. The natural function of AcrAB is presumed to be the pumping out of bile salts and their derivatives to survive in the presence of the high concentrations of these detergents in the natural habitat of *E. coli*. AcrAB catalyses efflux driven by proton motive force; this has been experimentally proved in intact cells using uncouplers and ionophores¹¹, and also in reconstituted proteoliposomes using artificial proton motive force⁹.

AcrB is the active part of the AcrAB–TolC drug export complex. It is composed of 1,049 amino-acid residues that form twelve putative

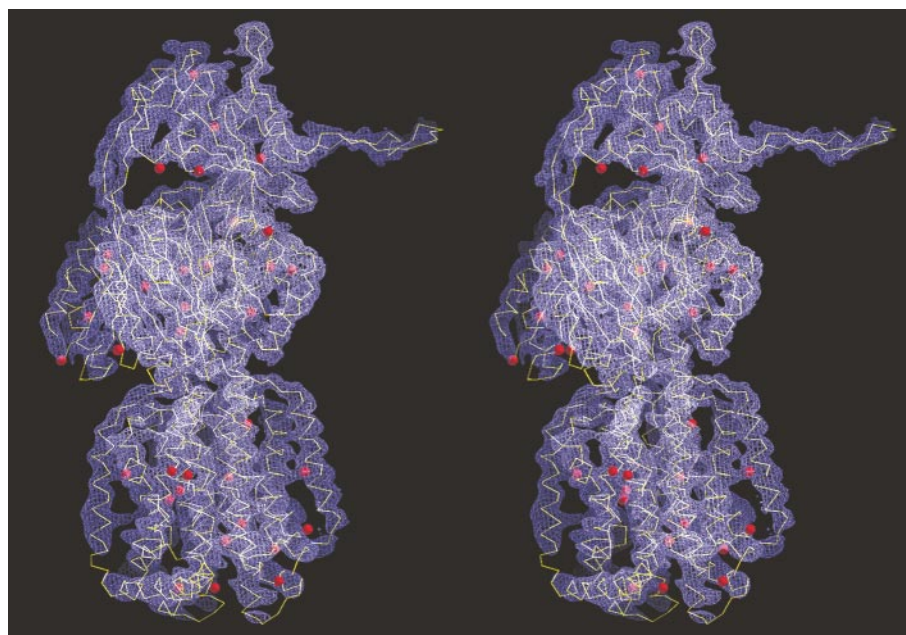


Figure 1 Stereo view of the experimental electron density map (2.0 σ contoured, purple) of the AcrB protomer at 3.5 Å resolution. The α trace (yellow lines) is superimposed on the

electron density map. This figure was prepared with program O (ref. 41). Thirty-eight selenium atoms are depicted as red balls.

transmembrane α -helices and two large hydrophilic loops on the periplasmic surface¹². AcrA is an elongated molecule that is presumed to link or fuse the inner and outer membranes^{13,14}. When engaged with its export substrate, AcrAB recruits TolC to form a transient complex that actively transports the substrates across both membranes. As shown in the crystal structure of TolC¹⁵, it comprises a trimer with a characteristic ‘channel-tunnel’ structure composed of a porin-like outer-membrane channel 40 Å long connected with a cylinder¹⁵ 100 Å long protruding into the periplasm.

Homologues of AcrB are widely distributed not only in Gram-negative bacteria but also in Gram-positive ones^{5,16} and confer intrinsic drug tolerance to microorganisms. Also, mammalian cells have an RND-type transporter that plays an important physiological role. This protein, the Niemann–Pick C1, is responsible for the autosomal recessive disorder characterized by cholesterol accumulation in lysosomes. It has been identified as an RND permease that transports fatty acids from the endosome/lysosome system¹⁷.

Recently, the crystal structures of ion-translocating membrane proteins such as K⁺-channels¹⁸ and Ca²⁺-ATPases¹⁹ were determined. The crystal structures of the lipid flippase MsbA²⁰ and the vitamin-B12-transporting ABC transporter²¹ BtuCD of *E. coli* have also been reported. However, attempts to obtain high-resolution structures of the largest group of transport proteins, the proton-coupled secondary transporters, have been thus far unsuccessful. Electron microscopic analyses of several secondary transporters were recently reported^{22–25}; however, the resolution was too low to assign helix-packing or to understand the mechanistic aspects of secondary transport.

Here, we determined the X-ray crystal structure of a secondary transporter AcrB, which exports a wide variety of organic compounds from the cell interior and also from the periplasm into the medium in combination with MFP and TolC. The structure of AcrB will greatly improve our understanding of the fundamental mechanism underlying active solute transport in general and multidrug transport in particular.

Structure determination

Initial phases at 4.3 Å were determined by the multiple isomorphous replacement (MIR) method using mercury, platinum and osmium derivatives. Out of 42 methionine residues, 38 sites were identified in the selenium anomalous difference Fourier map. MIR phases calculated by including selenomethionine-replaced crystal were extended to a 3.5 Å resolution by solvent flattening²⁶. The backbone of the protein was successfully traced in the map as shown in Fig. 1 (see also Supplementary Fig. 1), and its amino-acid sequence was consistent with 38 selenomethionine sites and bulky electron-density contours of aromatic residues. A total of 1,006 amino acids out of 1,049 residues were located in the 3.5 Å map. Structure refinement by the program CNS reduced an *R* and *R*_{free} to 0.290 and 0.355, respectively. The main-chain dihedral angles for 70.6% of the non-glycine residues were in the most favoured region and only 0.5% of the non-glycine residues were in the disallowed regions of the Ramachandran plots²⁷. The structural analysis is summarized in Table 1.

Overall architecture

AcrB comprises a trimer of 1,049-residue protomers (Fig. 2). The appearance of the trimer is that of a jellyfish with a three-fold symmetry axis perpendicular to the membrane plane. It comprises an extra-membrane (periplasmic) headpiece and a transmembrane region approximately 70 Å and 50 Å thick, respectively (Fig. 2a; see also Supplementary Information movie 1). The boundary of the membrane-embedded region (depicted as dotted lines in Fig. 2a) was estimated by the hydrophobic surface of the protein and the distribution of the aromatic amino acids. The maximum diameters of the headpiece and transmembrane region are approximately 100 Å and 80 Å, respectively. The headpiece is divided into two stacked parts. The upper and lower parts are 30 Å and 40 Å thick, respectively. The side view of the upper part has a trapezoidal appearance and is about 70 Å wide at the bottom and about 40 Å wide at the top. The crystal structure of TolC¹⁵ indicates that it protrudes 100 Å into the periplasm. The sum of the periplasmic

Table 1 Data collection and crystallographic analysis

	Native	Derivatives			
		Hg	Pt	Os	Se
Space group	<i>R</i> 32				
Wavelength (Å)	0.9000	0.9000	0.9000	0.9000	0.9795
Resolution (Å)	3.5 (3.63–3.5)	4.3 (4.45–4.3)	4.3 (4.45–4.3)	4.4 (4.56–4.4)	5.0 (5.1–5.0)
Observed reflections*	198,425 (13, 165)	102,662 (9,465)	87,755 (7,482)	96,176 (9,107)	131,081 (3,590)
Independent reflections	26,406 (2,584)	14,401 (1,406)	13,908 (1,393)	13,510 (1,338)	8,109 (720)
<i>I</i> / σ (<i>I</i>)	11.5 (2.6)	6.6 (5.4)	10 (6.5)	8.2 (5.0)	12.7 (6.9)
Averaged redundancy†	7.5 (5.1)	7.1 (6.7)	6.3 (5.3)	7.1 (6.8)	16.2 (5.0)
Completeness	98.8 (98.6)	99.7 (100)	94.9 (95.8)	99.3 (100)	87.9 (78.7)
<i>R</i> _{merge} ‡	9.0 (36.5)	12 (30)	10.8 (24.6)	11.1 (39.2)	14.2 (33.2)
<i>R</i> _{iso} §		14.5	17.2	28.9	16.9
<i>R</i> _{cullis}		0.81	0.86	0.83	0.96
Phasing power¶		0.53	0.43	0.54	0.28
Overall figure of merit#		0.32			
Refinement					
Resolution range (Å)	8.0–3.5				
<i>R</i> factor (%)	29.0				
<i>R</i> _{free} (%)	35.5				
Bond length (Å)	0.0032				
Bond angles (degrees)	0.867				
Average <i>B</i> -factor for main-chain atoms (Å ²)	95				
Average <i>B</i> -factor for side-chain atoms (Å ²)	117				

All diffraction data used in this structural analysis were collected on BL44XU at SPring-8. All data were collected at 100 K.
* The numbers in parentheses are for the highest-resolution shells.
† Redundancy is the number of observed reflections for each independent reflection.
‡ $R_{\text{merge}} = \sum_i \sum_h |I(h,i) - \langle I(h) \rangle| / \sum_i \sum_h I(h,i)$, where $I(h,i)$ is the intensity value of the *i*th measurement of *h*, and $\langle I(h) \rangle$ is the corresponding mean value of $I(h)$ for all *i* measurements; the summation is over the reflections with $I/\sigma(I)$ larger than 0.0.
§ $R_{\text{iso}} = \sum_h |F_{\text{PH}} - F_{\text{P}}| / \sum_h F_{\text{PH}}$, where F_{PH} and F_{P} are the derivative and the native structure factor amplitudes, respectively.
|| $R_{\text{cullis}} = \sum_h |F_{\text{PH}} - F_{\text{P}}| - F_{\text{H}}(\text{calc}) / \sum_h |F_{\text{PH}} - F_{\text{P}}|$, where $F_{\text{H}}(\text{calc})$ is the calculated heavy-atom structure factor. The summation is over the centric reflections only.
¶ Phasing power is root-mean-square (r.m.s.) isomorphous difference divided by r.m.s. residual lack of closure.
Figure of merit, $\langle |\sum P(\alpha) e^{i\alpha} / \sum P(\alpha)| \rangle$, where α is the phase and $P(\alpha)$ is the phase probability distribution.

length of AcrB and TolC is about 170 Å, which is enough to cross the periplasmic space. Given that the diameter of the top of the headpiece is almost equal to that of the TolC bottom¹⁵, AcrB and TolC might be directly docked with each other. Thus, the domains of protomers forming the upper part are tentatively named TolC docking domains.

Viewed from the top, the upper part is open like a funnel (Fig. 2b; see also Supplementary Information movie 2). The internal diameter of the funnel opening is about 30 Å, which is consistent with that of the TolC conduit¹⁵. Three α-helices form a pore at the centre of the headpiece (Fig. 2b). Each protomer contributes one pore helix. The pore exists in the lower part of the headpiece and connects with the bottom of the funnel. The pore seems to be closed in the crystal structure, probably because it does not contain a bound substrate. The long hairpin structure (~35 Å) protrudes from the TolC docking domain of every protomer and inserts into that of the next protomer. This interlocking arrangement presumably provides a very strong interaction holding the headpiece together.

In contrast to the headpiece, the transmembrane domains of the

protomers seem to be loosely packed. Three transmembrane domains of the protomers are arranged in a ring-like manner with a central hole (Fig. 2c). This hole crosses the membrane and extends to the bottom of the headpiece. The diameter of this hole is about 30 Å. It may be hard to keep phospholipids out of such a large transmembrane hole surrounded by the loosely interacting protomers. Thus, the transmembrane part of this hole might be filled with phospholipids.

Transmembrane domain structure

The transmembrane domain of each protomer contains twelve transmembrane α-helices (Fig. 2c). The number of the transmembrane segments is consistent with the hydropathy plot-based prediction and topology studies by site-directed chemical modification of cysteine mutants¹². A pseudo-two-fold symmetry axis exists in each transmembrane domain, that is, the six N-terminal helices are symmetrically arranged with the six C-terminal helices. The root mean square (r.m.s.) difference value of the superimposition of Co traces of the amino and carboxy-terminal halves of the transmem-

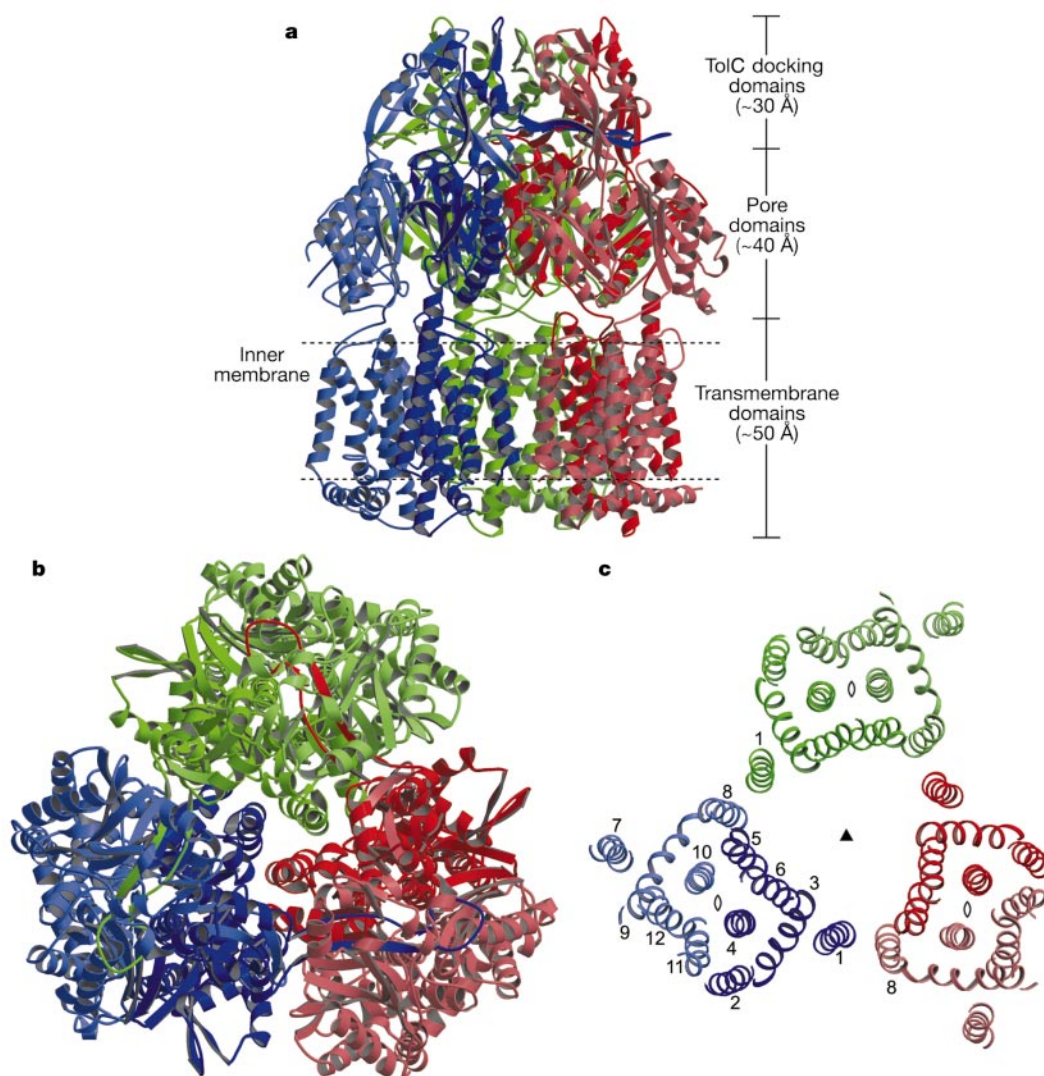


Figure 2 Structure of AcrB. **a**, Side view of a ribbon representation. Three protomers are individually coloured (blue, green and red). The N-terminal and C-terminal halves of the protomers are depicted as dark and pale colours, respectively. The extra-membrane (periplasmic) headpiece (TolC docking domains and pore domains) is at the top, and the transmembrane region is at the bottom. **b**, Top view of a ribbon representation. The protomers are individually coloured as in **a**. **c**, Structure within a slab (~23 Å) of the

transmembrane domain parallel to the membrane plane near the periplasmic surface. The protomers are individually coloured as in **a** and **b**. Three-fold and pseudo-two-fold rotation axes are indicated. The label numbers indicate the transmembrane helix numbers (TMx in Fig. 3a). This figure, Fig. 3b and c, Fig. 4 and Fig. 5 were generated by MOLSCRIPT⁴³ and Raster3D⁴⁴.

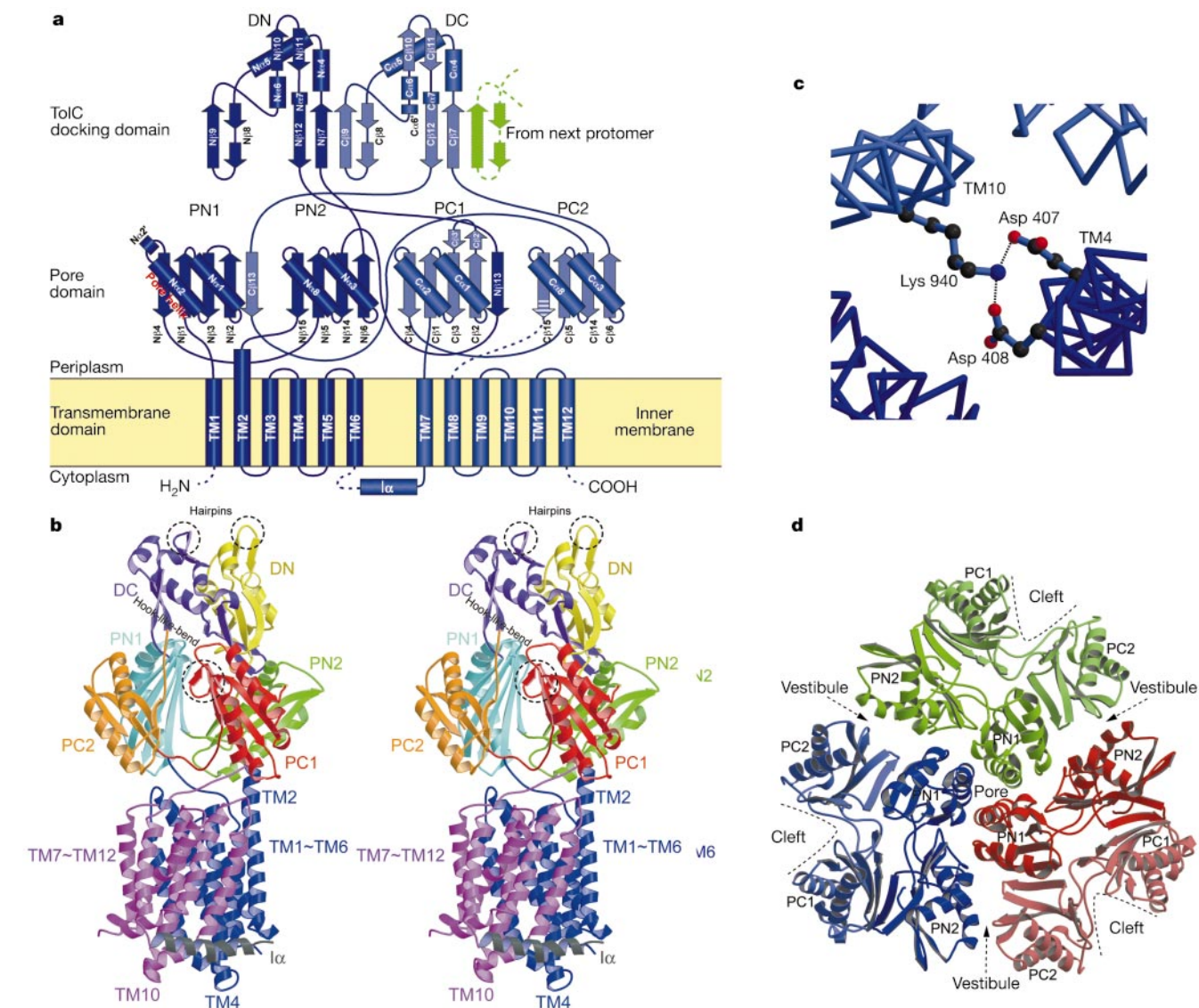


Figure 3 The structure of a single protomer. **a**, Topology diagram of the protomer. Secondary structure elements are indicated: TM, transmembrane helices; N α and N β , helices and strands, respectively, in the N-terminal half of the headpiece; C α and C β , helices and strands, respectively, in the C-terminal half of the headpiece; I α , a helix attached to the cytoplasmic surface of the membrane. N- and C-terminal halves are depicted in dark and pale blue, respectively. Dotted lines depict disordered polypeptide segments, positions 1–6, 499–512, 711, 860–868 and 1,037–1,049. The pore domain (lower part of the headpiece) is divided into four subdomains, PN1, PN2, PC1 and PC2. The TolC-docking domain (upper part of the headpiece) is divided into two subdomains, DN and DC. The hairpin structure inserted into the TolC-docking domain from the next protomer is depicted in green. **b**, A stereo view of ribbon representation of a protomer

viewed from outside the headpiece. Subdomains are depicted in different colours: PN1, cyan; PN2, green; PC1, red; PC2, orange; DN, yellow; DC, purple; N-terminal TM2, blue; C-terminal TM, magenta; I α , grey. There is a cleft perpendicular to the membrane plane between PC1 and PC2. A short hairpin structure at the depth of the cleft is labelled 'Hook-like bend'. Two vertical hairpins at the top of DN and DC are depicted in dotted circles. **c**, Ion pairs between Asp 407, Asp 408 and Lys 940 in TM4 and TM10 at the view from the cytoplasmic side. Side chains of Asp 407, Asp 408 and Lys 940 are shown in ball-and-stick representation, with the positions of the C, N and O atoms indicated by grey, blue and red balls, respectively. **d**, Cut view of the pore domains from the boundary between the pore and TolC-docking domains.

brane domain is ~ 2.4 Å. The membrane domain contains an additional extra-membrane α -helix (I α) located between TM6 and TM7 attached to the cytoplasmic membrane surface (Fig. 3a and b).

The inter-protomer interaction in the transmembrane region is restricted to the surface buried between TM1 and TM8 (Fig. 2c). TM4 and TM10 form the centre of the transmembrane helix bundle (Fig. 2c). These helices are long and protrude beyond the cytoplasmic surface of the membrane (Fig. 3b). TM2 is also a long helix that protrudes upwards across the boundary of the membrane (Fig. 3b). TM8 corresponds to TM2 in the C-terminal half. However, the polypeptide segment of the top of TM8 is disordered (amino-acid

residues 860–868). Except for the asymmetry of the top of TM2 and TM8, pseudo-two-fold symmetry of the N- and C-terminal six-helix bundles can be observed in Fig. 2c.

According to the results of site-directed mutagenesis studies on MexB²⁸ and AcrB (A. Saito, N. Tamura, T. Hirata, S.M. and A.Y., unpublished work), the transmembrane region contains three functionally essential charged residues: Asp 407, Asp 480 and Lys 940. When these residues are replaced with any other amino-acid residues, the resulting mutants completely lose the drug resistance. These residues are located in the middle of TM4 and TM10, and form ion pairs (Fig. 3c). These residues are possible candidates for the proton-translocating pathway.

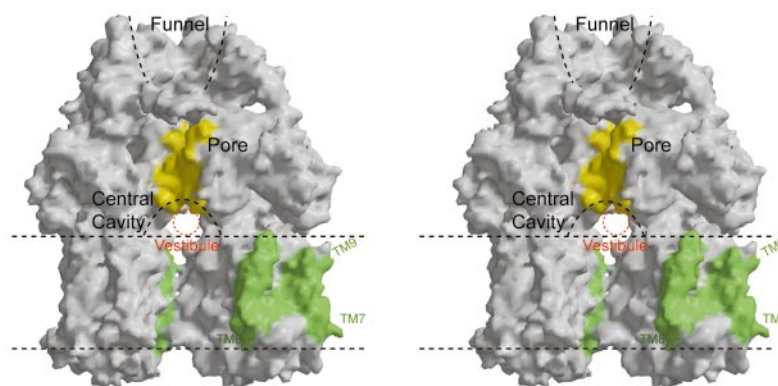


Figure 4 A cutaway stereo view displaying the solvent-accessible surface of AcrB rendered with GRASP⁴⁵. The front protomer is removed. A hairpin inserted into the removed protomer is viewed at the upper left as a stick-like protuberance. A hole for insertion of a hairpin from the removed protomer is also viewed at the upper right. The

yellow areas of the surface are coloured according to residues from Asp 99 to Leu 118 in pore helices (N α 2). The pale green areas are coloured according to residues in TM7 (Gly 539–Val 557) and TM8–TM9 (Ser 869–Phe 918). Frameworks of the funnel and the cavity are indicated by dotted lines.

Pore domain structure

The pore domain is composed of four subdomains: PN1, PN2, PC1 and PC2 (Fig. 3a and b). PN1 and PN2 comprise the polypeptide segment between TM1 and TM2, and PC1 and PC2 comprise the segment between TM7 and TM8. All of these subdomains contain a characteristic structural motif. That is, two β -strand– α -helix– β -strand motifs are directly repeated and sandwiched with each other. This motif forms a structure in which two α -helices are located on a four-stranded antiparallel β -sheet. As found by a database search (Dali/FSSP²⁹), this motif also exists in carboxypeptidase G2 (ref. 30). In PN1 and PC1, there is an additional antiparallel β -strand from the other half of the protomer (Fig. 3a). Thus, the antiparallel β -sheets in PN1 and PC1 are composed of five β -strands. On the other hand, in PN2 and PC2, the repeat of the β – α – β motifs is interrupted by switching to the TolC docking domain, and PC1 and PN1, respectively (Fig. 3a). The exception of the topological symmetry in the pore domain is a short hairpin, C β 2'–C β 3', and a short α -helix, N α 2'. The pore helix is an N α 2 in PN1 (Fig. 3a).

Four subdomains in the pore domain are packed with their sheets back to back in the centre, placing the α -helices on the outside (Fig. 3b). The deep cleft perpendicular to the membrane plane exists between PC1 and PC2. Its length is approximately 40 Å, its width 20 Å and its depth 15 Å. The vertical cleft along to the outer surface of the headpiece might be a binding site for AcrA, which is estimated to have an elongated shape¹³. At the depth of the cleft, there is a short hairpin structure, C β 2'–C β 3', in PC1, which forms a hook-like bend.

The lower part of the headpiece is composed of three pore domains, one for every protomer (Fig. 3d). The three pore domains combine to form a central pore in the headpiece. Viewed from the top, PN1, PN2, PC1 and PC2 are arranged in a clockwise order. PN1 is inside, and PN2, PC1 and PC2 are outside the headpiece (Fig. 3d). The pore-forming helix is the second helix (N α 2) of PN1. A cleft exists between PC1 and PC2 at the periphery of each protomer (Fig. 3d). Vestibules are open at the side of the headpiece between PN2 and PC2. They lead to a cavity located at the bottom of the pore as described below (Fig. 4).

TolC docking domain structure

The TolC docking domain is composed of two subdomains, DN and DC (Fig. 3a). Each subdomain contains a four-stranded mixed β -sheet. Two antiparallel β -strands are placed parallel to a hairpin structure from the other half or from the other protomer. The long hairpin structure (~35 Å) protrudes like a beak from subdomain DN (Fig. 3a; see also Supplementary Fig. 2). A vertical hairpin structure is located at the top of each subdomain of the TolC docking domain (Fig. 3a and b).

Three TolC docking domains form a funnel-like structure (Fig. 2b; see also Supplementary Information movie 2). The diameter of the top of the TolC docking domain is about the same as that of the bottom of TolC. The top of AcrB and the bottom of TolC fit well with each other by manual docking (Fig. 5; see also Supplementary Fig. 3). Six vertical hairpins, in total, at the top of AcrB trimer can contact with the six α -helix–turn– α -helix structures at the bottom of TolC¹⁵ to form a tight seal, by inspection.

Proposed transport mechanism

Figure 4 shows a one-protomer cutaway stereo view displaying the solvent-accessible surface of AcrB. A stick-like protrusion from the upper left of the headpiece is a hairpin structure inserted into the front protomer, which has been removed. The hole at the upper right of the headpiece is for insertion of the hairpin structure from the removed protomer. The funnel opens widely at the top of the headpiece. The bottom end of the funnel is narrow and connects

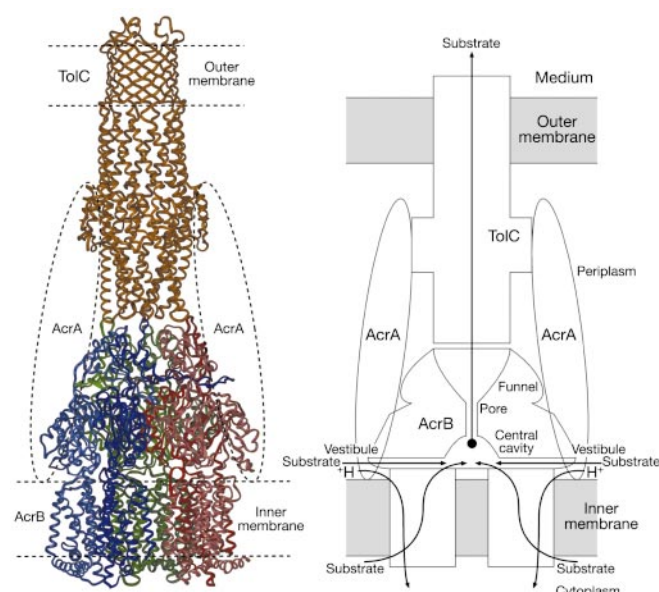


Figure 5 Proposed model of the AcrB–AcrA–TolC complex and the schematic mechanism of multidrug export mediated by AcrAB–TolC system. TolC structure¹⁵ is manually docked to AcrB. Dotted ovals indicate AcrA molecules. This figure was produced using Insight II/Discover (Biosym/MSI).

with the pore. The proximal end of the pore connects a cavity which is an extramembrane part of the central transmembrane hole. If the hole is filled with phospholipids, the depth of the cavity between the end of the pore and the membrane plane is about 15 Å. These are vestibules open from the cavity into the periplasm at the side of the headpiece near the membrane plane. The vestibules exist between every protomer. Because a front protomer is removed from the trimer in Fig. 4, only one vestibule can be observed at the centre of the cavity, like a window open backward. Two other vestibules also exist in both sides of the cavity, although they are hardly distinguishable in Fig. 4. A substrate located on the membrane plane or in the outer leaflet of the membrane might gain access to the cavity through these vestibules.

There is a groove in the membrane-exposed peripheral surface of the transmembrane domain of each protomer between TM8 and TM7. It extends across the whole membrane-embedded surface (Fig. 4, green-coloured region). The bottom of the groove is formed by TM9. Because TM9 is tilted, the groove is shallow at the cytoplasmic end and deep at the periplasmic end. The top of this groove is connected with the cavity via the disordered region of the top of TM8. The disordered region is close to the vestibule. This groove might be a pathway for substrate translocation across the membrane. The shape of the groove (shallow at the cytoplasmic side and deep at the periplasmic side) suggests that the transport through this groove of substrates located in the inner leaflet of the membrane might be more favourable than those in the cytoplasm.

Our AcrB crystal structure provides a structural basis for considering the molecular mechanism of multidrug export. The structure is consistent with the possibility of a direct interaction between TolC and AcrB. When the substrate is transported, AcrB might recruit TolC to form a direct transit pathway from the cytoplasm to the extracellular milieu (Fig. 5). Two putative pathways for substrate entry into AcrB are seen (Fig. 5). Substrates located in the cytoplasm or inner leaflet of the membrane might be transported across the membrane through the transmembrane groove at the periphery of each transmembrane domain. Then, they might be collected in the central cavity. The possibility that substrates pass through the centre of each protomer cannot be ruled out, as the groove at the hydrophobic peripheral surface of the transmembrane domain seems to be favourable for amphiphilic substrates partitioned into the inner leaflet of the membrane. There is also a possibility that the trimer centre hole plays some role in the substrate translocation. However, this hole might be filled with phospholipids. Substrates located on the outer surface or in the outer leaflet of the membrane could gain access to the cavity through the vestibules open into the periplasm. They would then be actively transported through the pore into the TolC tunnel via the funnel. The ion pairs between Lys 940, Asp 407 and Asp 408 are possible candidates for the transmembrane proton translocation site. When these aspartic acids are protonated, the ion pairs are disrupted. The helices TM4 and TM10 may then undergo a conformational change. This conformational change may be transduced to the pore region by a remote conformational coupling and thereby open the pore.

The AcrAB–TolC system exports a wide variety of structurally unrelated compounds, most of which are amphiphilic^{1,4}. In particular, the unique property of this system and other MFP-dependent systems, such as the MexAB–OprM in *Pseudomonas aeruginosa*, is its ability to confer resistance to β -lactam antibiotics^{31,32}. The targets of the β -lactam antibiotics are periplasmic enzymes that participate in cell wall biosynthesis, and secondary transporters usually do not export solutes from the periplasm, so it has been surprising that RND transporters confer resistance to these drugs.

How do RND transporters actively export drugs from the periplasm? Nikaido *et al.* proposed a dual entrance model for the AcrB export mechanism; that is, substrates are taken up from the inner and outer leaflets of the lipid bilayer of the cell membrane³³.

However, there has been no experimental evidence for this model. In this regard, the crystal structure of AcrB provides a possible explanation. As shown, AcrB contains vestibules open to the periplasm at the side of the headpiece. They permit direct access of the substrates from the periplasm or outer leaflet of the membrane. AcrB can also mediate the translocation of the substrates from the cytoplasm or the inner leaflet of the membrane. This structure supports the dual-entrance model for AcrB-mediated multidrug export. □

Methods

Protein preparation

A histidine-tagged AcrB-overproducing plasmid pAcBH was previously constructed¹². The AcrB was overproduced in *E. coli* JM109. The cells were disrupted with Microfluidizer M-110EH (Microfluidics Corp.) and the membrane fractions were collected and washed with several ultracentrifugation steps at 150,000g for 90 min. Purified membranes were resuspended in buffered glycerol (50 mM Tris-HCl, pH 7.0, 10% glycerol). Then membrane proteins were solubilized in 2% n-dodecyl- β -D-maltoside (DDM) (Anatrace). Lipids and debris were removed by ultracentrifugation at 170,000g for 60 min. Extracted histidine-tagged AcrB was purified with affinity column chromatography using Chelating Sepharose (Amersham Bioscience) immobilized with Ni²⁺ equilibrated with buffer (20 mM Tris-HCl, pH 7.5, 0.3 M NaCl, 10% glycerol and 0.2% DDM). The column was washed using 25 mM imidazole added to the above buffer and then 100 mM imidazole. Purified AcrB was eluted with 300 mM imidazole. Imidazole was removed by concentration-dilution steps of proteins three times on a ultrafiltration membrane. Proteins were finally concentrated to 35–40 mg ml⁻¹. For proteins labelled with seleno-L-methionine, transformed *E. coli* B834DE3 (Novagen, WI, USA) were grown in minimal medium (L-methionine was replaced by seleno-L-methionine), and the protein was purified as above. Replacement of methionine sulphur with selenium was confirmed by X-ray fluorescence analysis on BL44XU at SPring-8.

Preparation of native and derivative crystals

Crystals were grown by vapour diffusion from sitting drops at 25 °C. A protein solution containing 28 mg ml⁻¹ histidine-tagged AcrB protein, 20 mM sodium phosphate (pH 6.2), 10% glycerol and 0.2% dodecylmaltoside was mixed (1:1) with a reservoir solution containing 15–16% polyethylene glycol 2000, 80 mM sodium phosphate (pH 6.2), 20 mM sodium citrate-HCl (pH 5.6). The crystals were grown within a few days to optimal size (0.3 × 0.3 × 0.3 mm³). The concentration of glycerol was gradually increased from 5% to 30% in 5% steps by the soaking method for optimal cryo-protection. Crystals were picked up using nylon loops (Hampton Research) for flash-cooling in cold nitrogen gas from a cryostat (Rigaku, Japan) For derivatization, crystals were soaked for 1 h in cryo-protection buffer containing 0.5 mM K₂HgI₄, 1 mM K₂PtCl₆ and 1 mM OsCl₃ before freezing.

Crystallographic analysis

All data sets were collected on BL44XU at SPring-8 with imaging-plate detector DIP2040 (MAC-Science) at cryogenic temperature (100 K). The diffraction data were processed and scaled with DENZO and SCALEPACK³⁴, respectively. The native crystal belongs to the space group R32 with cell dimensions of $a = b = 144.3$ Å, $c = 519.0$ Å in the hexagonal setting. Three heavy-atom derivatives of mercury, osmium and platinum, and the selenomethionine-replaced crystal were isomorphous with the native crystal. Initial phases were first calculated by program SOLVE³⁵ derived from the mercury derivative. The heavy-atom sites of the platinum and osmium derivatives were determined in difference Fourier maps calculated with the single isomorphous phases of the mercury derivative. MIR phase refinement with three heavy-atom derivatives was performed at 4.3 Å resolution with MLPHARE^{36,37}. Anomalous dispersion effects of the heavy metals were taken into account for the phase estimation. To locate the methionine residues in the crystal, anomalous difference Fourier maps were calculated with coefficients of $(F_{sc}^+ - F_{sc}^-) \exp[i(\alpha_p - \pi/2)]$ at 5 Å resolution, where $(F_{sc}^+ - F_{sc}^-)$ was the Bijvoet difference of the selenomethionine-replaced crystal and where α_p was the MIR phase of the native crystal determined with the three derivatives. Out of 42 methionine sites of the protein molecule, 35 positions were identified by selenium peaks higher than 4 σ , and 38 sites were higher than 3 σ in the anomalous difference Fourier map. The selenomethionine-replaced crystal was included in MIR phases at 4.3 Å resolution. The phases were improved and extended to 3.5 Å resolution with RESOLVE^{38,39}. Model building was performed using program O (ref. 40) and model refinement was conducted using Crystallographic and NMR system (CNS)⁴¹. Out of 1,049 residues of the whole molecule, the atomic parameters of residues 7–498, 513–710, 712–859 and 869–1,036 converged well during the refinement. A crystallographic R and an R_{free} for 5% reflections excluded from the refinement were reduced to 0.290 and 0.355, respectively. The refined structure was validated using the program PROCHECK⁴². The main-chain dihedral angles for 70.6% of the non-glycine residues were in the most favoured regions of the Ramachandran plots²⁷; 26.3% in allowed regions; 2.6% in generously allowed regions; and 0.5% in the disallowed regions.

Received 17 May; accepted 6 August 2002; doi:10.1038/nature01050.

- Okusu, H., Ma, D. & Nikaido, H. AcrAB efflux pump plays a major role in the antibiotic resistance phenotype of *Escherichia coli* multiple-antibiotic-resistance (Mar) mutants. *J. Bacteriol.* **178**, 306–308 (1996).
- Ma, D., Cook, D. N., Hearst, J. E. & Nikaido, H. Efflux pumps and drug resistance in Gram-negative bacteria. *Trends Microbiol.* **2**, 489–493 (1994).

3. Paulsen, I. T., Nguyen, L., Sliwinski, M. K., Rabus, R. & Saier, M. H. Microbial genome analyses: comparative transport capabilities in eighteen prokaryotes. *J. Mol. Biol.* **301**, 75–100 (2000).
4. Sulavik, M. C. *et al.* Antibiotic susceptibility profiles of *Escherichia coli* strains lacking multidrug efflux pump genes. *Antimicrob. Agents Chemother.* **45**, 1126–1136 (2001).
5. Tseng, T.-T. *et al.* The RND permease superfamily: An ancient, ubiquitous and diverse family that includes human disease and development proteins. *J. Mol. Microbiol. Biotechnol.* **1**, 107–125 (1999).
6. Dinh, T., Paulsen, I. T. & Saier, M. H. Jr A family of extracytoplasmic proteins that allow transport of large molecules across the outer membrane of Gram-negative bacteria. *J. Bacteriol.* **176**, 3825–3831 (1994).
7. Paulsen, I. T., Park, J. H., Choi, P. S. & Saier, M. H. A family of Gram negative bacterial outer membrane factors that function in the export of proteins, carbohydrates, drugs, and heavy metals. *FEMS Microbiol. Lett.* **156**, 1–8 (1997).
8. Nikaido, H. How do exported proteins and antibiotics bypass the periplasm in Gram-negative bacterial cells? *Trends Microbiol.* **8**, 481–483 (2000).
9. Zgurskaya, H. I. & Nikaido, H. Bypassing the periplasm: Reconstitution of the AcrAB multidrug efflux pump of *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **96**, 7190–7195 (1999).
10. Nikaido, H. & Zgurskaya, H. I. AcrAB and related multidrug efflux pumps of *Escherichia coli*. *J. Mol. Microbiol. Biotechnol.* **3**, 215–218 (2001).
11. Thanassi, D. G., Cheng, L. W. & Nikaido, H. Active efflux of bile salts by *Escherichia coli*. *J. Bacteriol.* **179**, 2512–2518 (1997).
12. Fujihira, E., Tamura, N. & Yamaguchi, A. Membrane topology of a multidrug efflux transporter, AcrB, in *Escherichia coli*. *J. Biochem.* **131**, 145–151 (2002).
13. Zgurskaya, H. I. & Nikaido, H. AcrA is a highly asymmetric protein capable of spanning the periplasm. *J. Mol. Biol.* **285**, 409–420 (1999).
14. Avila-Sakar, A. J. *et al.* Lipid-layer crystallization and preliminary three-dimensional structural analysis of AcrA, the periplasmic component of a bacterial multidrug efflux pump. *J. Struct. Biol.* **136**, 81–88 (2001).
15. Koronakis, V., Shaff, A., Koronakis, E., Luisi, B. & Hughes, C. Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export. *Nature* **405**, 914–919 (2000).
16. Poole, K. Multidrug efflux pumps and antimicrobial resistance in *Pseudomonas aeruginosa* and related organisms. *J. Mol. Microbiol. Biotechnol.* **3**, 255–264 (2001).
17. Davis, J. P., Chen, F. W. & Ioannou, Y. A. Transmembrane molecular pump activity of Niemann-Pick C1 protein. *Science* **290**, 2295–2298 (2000).
18. Doyle, D. A. *et al.* The structure of the potassium channel: molecular basis of K⁺ conduction and selectivity. *Science* **280**, 69–77 (1998).
19. Toyoshima, T., Nakasako, M., Nomura, H. & Ogawa, H. Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 Å resolution. *Nature* **405**, 647–655 (2000).
20. Chang, G. & Roth, C. B. Structure of MsbA from *E. coli*: A homolog of the multidrug resistance ATP binding cassette (ABC) transporters. *Science* **293**, 1793–1800 (2001).
21. Locher, K. P., Lee, A. T. & Rees, D. C. The *E. coli* BtuCD structure: A framework for ABC transporter architecture and mechanism. *Science* **296**, 1091–1098 (2002).
22. Williams, K. A. Three-dimensional structure of the ion-coupled transport protein NhaA. *Nature* **403**, 112–115 (2000).
23. Zhuang, J. *et al.* Two-dimensional crystallization of *Escherichia coli* lactose permease. *J. Struct. Biol.* **125**, 63–75 (1999).
24. Yin, C.-C. *et al.* The quaternary molecular architecture of TetA, a secondary tetracycline transporter from *Escherichia coli*. *Mol. Microbiol.* **38**, 482–492 (2000).
25. Heymann, J. A. W. *et al.* Projection structure and molecular architecture of OxlT, a bacterial membrane transporter. *EMBO J.* **20**, 4408–4413 (2001).
26. Wang, B. C. Resolution of phase ambiguity in macromolecular crystallography. *Methods Enzymol.* **115**, 90–112 (1985).
27. Ramakrishnan, C. & Ramachandran, G. N. Stereochemical criteria for polypeptide and protein chain conformations. II. Allowed conformations for a pair of peptide units. *Biophys. J.* **5**, 909–933 (1965).
28. Guan, L. & Nakae, T. Identification of essential charged residues in transmembrane segments of the multidrug transporter MexB of *Pseudomonas aeruginosa*. *J. Bacteriol.* **183**, 1734–1739 (2001).
29. Holm, L. & Sander, C. Mapping the protein universe. *Science* **273**, 595–602 (1996).
30. Rowsell, S. *et al.* Crystal structure of carboxypeptidase G2, a bacterial enzyme with applications in cancer therapy. *Structure* **5**, 337–347 (1997).
31. Srikumar, R., Li, X. Z. & Poole, K. Inner membrane efflux components are responsible for beta-lactam specificity of multidrug efflux pumps in *Pseudomonas aeruginosa*. *J. Bacteriol.* **179**, 7875–7881 (1997).
32. Mazzariol, A., Cornaglia, G. & Nikaido, H. Contributions of the AmpC beta-lactamase and the AcrAB multidrug efflux system in intrinsic resistance of *Escherichia coli* K-12 to beta-lactams. *Antimicrob. Agents Chemother.* **44**, 1387–1390 (2000).
33. Nikaido, H., Basina, M., Nguyen, V. Y. & Rosenberg, E. Y. Multidrug efflux pump AcrAB of *Salmonella typhimurium* excretes only those β-lactam antibiotics containing lipophilic side chains. *J. Bacteriol.* **180**, 4686–4692 (1998).
34. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–325 (1997).
35. Terwilliger, T. C. & Berendzen, J. Automated structure solution for MIR and MAD. *Acta Crystallogr. D* **55**, 849–861 (1999).
36. Otwinowski, Z. MLPHARE, CCP4 Proc. 80 (Daresbury Laboratory, Warrington, UK, 1991).
37. Collaborative Computational Project No. 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
38. Terwilliger, T. C. Maximum-likelihood density modification. *Acta Crystallogr. D* **56**, 965–972 (2000).
39. Terwilliger, T. C. Reciprocal-space solvent flattening. *Acta Crystallogr. D* **55**, 1863–1871 (1999).
40. Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for binding protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
41. Brünger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
42. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemistry of protein structures. *J. Appl. Crystallogr.* **26**, 283–291 (1993).
43. Kraulis, P. J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
44. Merritt, E. A. & Bacon, D. J. Raster3D: photorealistic molecular graphics. *Methods Enzymol.* **277**, 505–524 (1997).
45. Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins Struct. Funct. Genet.* **11**, 281–296 (1991).

Supplementary Information accompanies the paper on *Nature's* website
(<http://www.nature.com>).

Acknowledgements We thank T. Tsukihara for advice on all stages of our X-ray crystallographic analysis and critical reading of this manuscript, and M. Yoshida, H. Nikaido and E. Bibi for critical reading and discussions. We are also grateful to A. Nakagawa, M. Kusunoki and G. Kurisu for help in the diffraction data collection and computation. We thank D. Goodhew-Davison for English correction. This work was supported by Grants-in-Aid from the Ministry of Education, Culture, Sports, Science and Technology of Japan. This work was done with the approval of the Joint Research Committee of Institute for Protein Research, Osaka University.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A. Y. (e-mail: akihito@sanken.osaka-u.ac.jp). Coordinates have been deposited with the Protein Data Bank under accession code 1IWG.

Indistinguishable photons from a single-photon device

Charles Santori*, David Fattal*, Jelena Vučković*, Glenn S. Solomon*† & Yoshihisa Yamamoto*‡

* Quantum Entanglement Project, ICORP, JST, E. L. Ginzton Laboratory, Stanford University, Stanford, California 94305-4088, USA

† Solid-State Photonics Laboratory, Stanford University, Stanford, California 94305-4085, USA

‡ NTT Basic Research Laboratories, Atsugi, Kanagawa, 243-0198, Japan

Single-photon sources have recently been demonstrated using a variety of devices, including molecules^{1–3}, mesoscopic quantum wells⁴, colour centres⁵, trapped ions⁶ and semiconductor quantum dots^{7–11}. Compared with a Poisson-distributed source of the same intensity, these sources rarely emit two or more photons in the same pulse. Numerous applications for single-photon sources have been proposed in the field of quantum information, but most—including linear-optical quantum computation¹²—also require consecutive photons to have identical wave packets. For a source based on a single quantum emitter, the emitter must therefore be excited in a rapid or deterministic way, and interact little with its surrounding environment. Here we test the indistinguishability of photons emitted by a semiconductor quantum dot in a microcavity through a Hong–Ou–Mandel-type two-photon interference experiment^{13,14}. We find that consecutive photons are largely indistinguishable, with a mean wave-packet overlap as large as 0.81, making this source useful in a variety of experiments in quantum optics and quantum information.

When identical single photons enter a 50–50 beam splitter from opposite sides, quantum mechanics predicts that both photons must leave in the same direction, if their wave packets overlap perfectly. This two-photon interference effect originates from the Bose–Einstein statistics of photons. This bunching effect was first observed using pairs of highly correlated photons produced by parametric downconversion¹⁴, but it should also occur with single, independently generated photons. Most proposed applications for single-photon sources in the field of quantum information (with the notable exception of quantum cryptography¹⁵) involve two-photon interference. Such applications include quantum teleportation¹⁶, post-selective production of polarization-entangled photons¹⁷, and linear-optics quantum computation¹². It is therefore important to demonstrate that consecutive photons emitted by a single-photon source are identical and exhibit mutual two-photon interference effects.

The experiment described here used a semiconductor quantum dot as the photon source. Quantum dots are attractive as single-photon sources because they are relatively stable, have narrow spectral linewidths and rapid radiative decay rates, and can be integrated into larger fabricated structures—such as microcavities—to improve the collection efficiency^{18,19}. A quantum dot excited on resonance by a pulsed source can have an extremely small probability of generating two photons in the same pulse—as required for this experiment. Furthermore, recent reports have indicated coherence times²⁰, and even time-averaged linewidths^{21,22}, fairly close to the radiative limit in some cases, suggesting that dephasing is slow, and thus indistinguishable photons may be achievable.

Our sample contains self-assembled InAs quantum dots (about $25\text{ }\mu\text{m}^{-2}$) embedded in GaAs and sandwiched between distributed-Bragg-reflector (DBR) mirrors, grown by molecular-beam epitaxy¹⁹. Pillars (Fig. 1a) with diameters ranging from 0.3 to $5\text{ }\mu\text{m}$ and heights of $5\text{ }\mu\text{m}$ were fabricated in a random distribution by chemically assisted ion beam etching (CAIBE), using sapphire dust particles as etch masks. The resulting microcavities, exhibiting

three-dimensional photon confinement, have quality factors of approximately 1,000 and measured spontaneous-emission rate enhancement (Purcell) factors as high as 5. Many pillars with only one or two quantum dots on resonance with a fundamental cavity mode were found. The sample was cooled to 3–7 K in a cryostat. To generate single photons, we focused 3-ps pulses from a Ti-sapphire laser every 13 ns onto these pillars from a steep angle. The laser was tuned to an excited-state absorption resonance of the quantum dot, typically 20–30 nm shorter in wavelength than the first-level emission wavelength. The quantum-dot emission was collected, and a single polarization was selected. The emission was then spectrally filtered with a resolution of about 0.1 nm using a diffraction grating, and coupled into a single-mode fibre.

By this method, we obtained bright, single-photon sources with excellent two-photon suppression and negligible background emission. We have chosen three quantum dots for this study, denoted as dots 1, 2 and 3, with emission wavelengths (in nm) of 931, 932 and 937, respectively. A photon-correlation measurement for dot 2 is shown in Fig. 1b. A parameter often used to quantify two-photon suppression is $g^{(2)}$, the probability of generating two photons in the same pulse, normalized by an equally bright Poisson-distributed source. We estimate $g^{(2)} = 0.053$, 0.067 and 0.071 for dots 1, 2 and 3, respectively. But for the experiment described below, the important parameter is the probability of generating two photons in the same pulse, for either of two consecutive pulses, divided by the

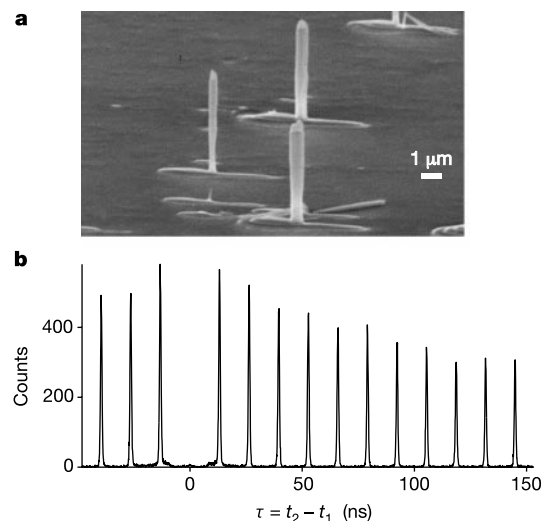


Figure 1 The single-photon source. **a**, Pillar microcavity structures containing InAs quantum dots in a one-wavelength-thick GaAs spacer, sandwiched between distributed-Bragg-reflector (DBR) mirrors, grown by molecular-beam epitaxy. The DBR mirrors were constructed by stacking quarter-wavelength-thick GaAs and AlAs layers on top of each other. There are 12 DBR pairs above, and 30 DBR pairs below, the spacer. Pillars with diameters ranging from 0.3 to $5\text{ }\mu\text{m}$ and heights of $5\text{ }\mu\text{m}$ were fabricated in a random distribution by chemically assisted ion beam etching (CAIBE) with Ar^+ ions and Cl_2 gas, using sapphire dust particles as etch masks. Owing to the irregular shapes of the pillars, the fundamental mode is typically polarization-nondegenerate. **b**, Photon correlation histogram of emission from quantum dot 2 under pulsed, resonant excitation, obtained using a Hanbury Brown and Twiss-type set-up. The emission was split into two paths by a beam splitter, each path leading to a photon counter. A histogram was generated of the relative delay $\tau = t_2 - t_1$ between a photon detection at one counter (t_1) and the other (t_2). The vanishing central peak is the signature of suppressed two-photon emission. The parameter g described in the text was obtained by dividing the area of the central peak at $\tau = 0$ by that of the nearest side peaks. The decrease of the side peaks away from $\tau = 0$ indicates blinking with a timescale of 85 ns, an effect that we usually see with resonant excitation⁸. For this measurement, the set-up shown in Fig. 3a was used, with one arm blocked.

probability of generating one photon in each pulse. We estimate this quantity to be $g = 0.039, 0.027$ and 0.025 for quantum dots 1, 2 and 3, respectively. The difference between $g^{(2)}$ and g is due to blinking in our source.

Two other properties of the quantum-dot emission are also important for the two-photon interference experiment described below: the spontaneous emission lifetime and the coherence length. The average emission intensity of quantum dots 1, 2 and 3 is plotted versus time after an excitation pulse, measured under resonant excitation by a streak camera (Fig. 2a). By fitting decaying exponential functions, we estimate the spontaneous emission lifetimes τ_s of dots 1, 2 and 3 to be (in ps) 89, 166 and 351, respectively. This variation is due largely to differences in how well each quantum dot couples to its microcavity. A Michelson interferometer is used to measure the coherence length of the time-averaged emission (Fig. 2b). The curves show how the interference fringe contrast varies with path-length difference, and give the magnitude of the Fourier transform of the intensity spectra. When we did not select a single polarization, we sometimes observed oscillatory behaviour due to polarization splitting of the emission lines²³. For dots 2 and 3 (with splittings of 13 and 17 μeV), we were able to eliminate this effect by selecting a particular linear polarization. For dot 1, the 45- μeV splitting could not easily be eliminated, probably because the quantum-dot emission couples to just one cavity mode having a polarization rotated $\sim 45^\circ$ relative to the splitting axis of the quantum dot. We estimate the $1/e$ coherence lengths τ_c (divided by c) for quantum dots 1, 2 and 3 to be (in ps) 48, 223 and 105, respectively. Quantum dot 2 is closest to being Fourier-transform-

limited, with $2\tau_s/\tau_c = 1.5$. When this ratio is equal to 1, no dephasing can be present, and perfect two-photon interference is expected.

The main elements of the two-photon interference experiment are shown in Fig. 3a. The single-photon source is as described above, except that the quantum dot is excited twice every 13 ns by a pair of equally intense pulses with 2 ns separation. Two pulses, each containing zero or one photons, emerge from the single-mode fibre. They are split into two arms by a beam splitter, with one arm ($2\text{ ns} + \Delta t$) longer than the other. The beams then recombine at a different place on the same beam splitter. The two outputs of this interferometer are collected by photon counters, and a photon correlation histogram is generated of the relative delay time $\tau = t_2 - t_1$ for two-photon coincidence events, where t_1 and t_2 are the times at which photons are detected at detectors 1 and 2, respectively. A histogram obtained in this way for dot 2 with $\Delta t = 0$ is shown in Fig. 3b.

Five peaks appear within the central cluster, corresponding to three types of coincidence events. For peaks 1 and 5 at $\tau = \mp 4\text{ ns}$, the first photon follows the short arm of the interferometer, the second photon follows the long arm, and one photon goes to each counter. For peaks 2 and 4 at $\tau = \mp 2\text{ ns}$, both photons follow the same arm. For peak 3 at $\tau = 0$, the first photon follows the long arm, and the second photon follows the short arm, so that the two photons collide upon their second pass through the beam splitter. Only in this case can two-photon interference occur, and for perfect two-photon interference, peak 3 vanishes.

When the source successfully delivers a pair of photons, the two-

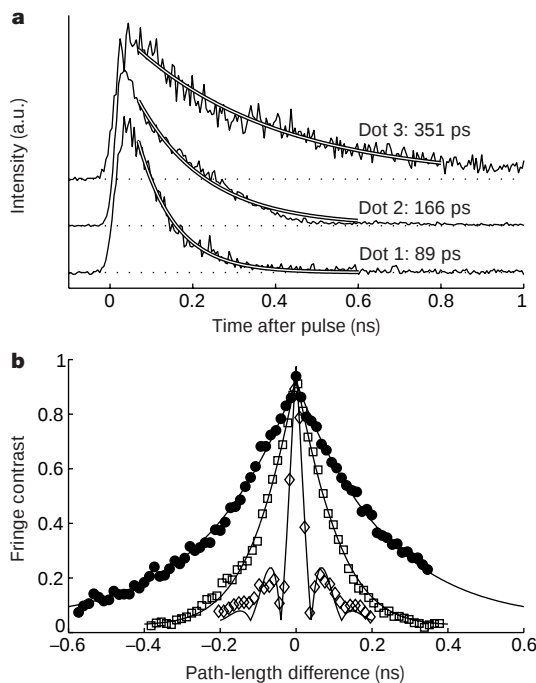


Figure 2 Time-averaged emission properties of quantum dots 1, 2 and 3. **a**, Spontaneous emission decay under resonant excitation, measured by a streak camera. By fitting exponentials, lifetimes τ_s of 89, 166 and 351 ps for dots 1, 2 and 3, respectively, are obtained. **b**, Coherence length, measured using a Michelson interferometer, showing fringe contrast versus path-length difference. The set-up was similar to the one shown in Fig. 3a, but with the 2-ns delay removed. The fringe contrast was measured by monitoring the intensity of one of the interferometer outputs while varying one of the arm lengths over several wavelengths using a piezoelectric transducer. The arm length was then moved over long distances by a motor stage. The $1/e$ coherence lengths τ_c are 48, 223 and 105 ps for dots 1 (diamonds), 2 (filled circles) and 3 (squares), respectively.

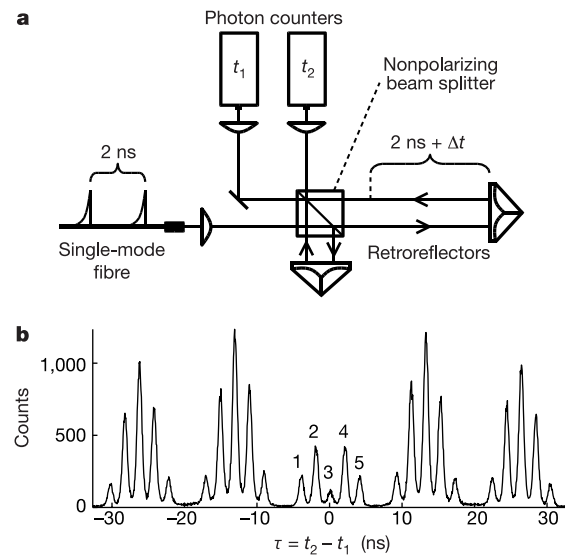


Figure 3 Two-photon interference experiment. **a**, Every 13 ns, two pulses, separated by 2 ns and containing 0 or 1 photons, arrive through a single-mode fibre. The pulses are interfered with each other using a Michelson-type interferometer with a ($2\text{ ns} + \Delta t$) path-length difference. Corner-cube retroreflectors are used at the ends of the arms, so that the mode overlap is insensitive to slight angular misalignment of the optical elements. The length of the short arm can be adjusted over long distances by a 15-cm motor stage. The fringe contrast measured using a laser with a long coherence length was 0.92, limited by optical surface imperfections. The interferometer outputs are collected by photon counters, and the resulting electronic signals are correlated using a time-to-amplitude converter followed by a multi-channel analyser card, which generates a histogram of the relative delay time $\tau = t_2 - t_1$ between a photon detection at one counter (t_1) and the other (t_2). **b**, Such a histogram (53-ps bin size) obtained for quantum dot 2, with $\Delta t = 0$. The number of repetitions was $N = 2.3 \times 10^{10}$ (5 min), and the combined two-photon generation and detection efficiency was $\eta^{(2)} = 2.5 \times 10^{-6}$, which includes all losses in the experimental set-up. The small area of peak 3 demonstrates two-photon interference.

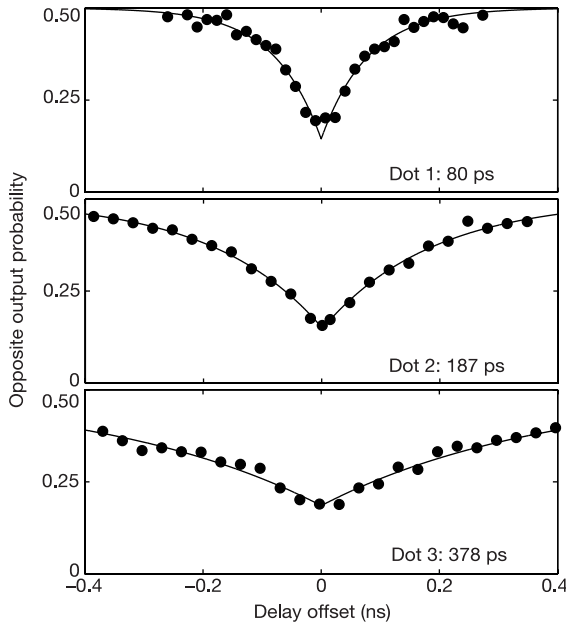


Figure 4 The probability that two photons that collide at the beam splitter leave in opposite directions, plotted as a function of interferometer delay offset, Δt . Data are shown for quantum dots 1, 2 and 3. The drops in the coincidence rate near zero offset demonstrate two-photon interference. From model fits (solid lines), the $1/e$ widths of the dips are estimated as 80, 187 and 378 ps for dots 1, 2 and 3, respectively, in close agreement with the measured spontaneous emission lifetimes.

photon state can be written as

$$|\psi\rangle = \int ds x(s) \int dt y(t) a^\dagger(s) a^\dagger(t + 2\text{ ns}) |0\rangle \quad (1)$$

where $a^\dagger(t)$ is the photon creation operator at time t , $x(s)$ and $y(t)$ define the photon wave packets, and $|0\rangle$ is the vacuum state. We assume that the photon wave packets are much shorter than 2 ns. In the limit of low collection efficiency, the mean areas of peaks 1–5 are

$$\begin{aligned} A_1 &= N\eta^{(2)}R^3T, \quad A_2 = N\eta^{(2)}[R^3T(1+2g) + RT^3] \\ A_3 &= N\eta^{(2)}[(R^3T + RT^3)(1+2g) - 2(1-\varepsilon)^2R^2T^2V(\Delta t)] \quad (2) \\ A_4 &= N\eta^{(2)}[R^3T + RT^3(1+2g)], \quad A_5 = N\eta^{(2)}RT^3 \end{aligned}$$

where N is the number of repetitions, $\eta^{(2)}$ is the combined two-photon generation and detection efficiency, and R and T are the beam-splitter intensity coefficients of reflection and transmission, respectively. As defined above, the parameter g characterizes the two-photon emission probability, with $g = 0$ for an ideal single-photon source, and $g = 1$ for a Poisson-distributed source (without blinking). The parameter $1 - \varepsilon$ is the interference fringe contrast measured when an ideal monochromatic calibration source is sent into the interferometer, and accounts for optical surface imperfections. The parameter $V(\Delta t) = \langle |\int dt x(t) y^*(t + \Delta t)|^2 \rangle$ in the expression for peak 3 is the mean overlap between the wave packets of the two photons for interferometer path-length difference ($2\text{ ns} + \Delta t$). An ensemble average is performed over all possible two-photon states generated by the source.

The signature of two-photon interference that we observe is the small size of peak 3 in Fig. 3b, compared with peaks 2 and 4. We define the quantity $M(\Delta t) = A_3/(A_2 + A_4)$ in terms of the peak areas in equation (2). This quantity is equal to the conditional probability, given that two photons collide at the beam splitter, that the photons leave in opposite directions, in the limit $g \approx 0$. We measured $M(\Delta t)$ while varying the interferometer path length offset

Table 1 Summary of quantum-dot parameters

	$g^{(2)}$	g	τ_s (ps)	τ_c (ps)	τ_m (ps)	$V(0)$
Dot 1	0.053	0.039	89	48	80	0.72
Dot 2	0.067	0.027	166	223	187	0.81
Dot 3	0.071	0.025	351	105	378	0.74

For the three quantum dots chosen for this study, we show the conventional two-photon suppression parameter $g^{(2)}$, the ratio g of the probability of emitting two photons in either of two consecutive pulses to the probability of emitting one photon in each pulse, the spontaneous emission lifetime τ_s , the coherence length τ_c , the $1/e$ width of the Mandel dip τ_m , and the two-photon overlap at zero path-length difference $V(0)$.

Δt (Fig. 4). For all three quantum dots, we observe reductions in the coincidence probability near $\Delta t = 0$, by factors of 0.61, 0.69 and 0.62 for dots 1, 2 and 3, respectively. The remaining coincidences we see are partly due to independently measured optical imperfections in our set-up, $R/T = 1.1$ and $(1 - \varepsilon) = 0.92$. Without these imperfections, the coincidence reduction factors would be $V(0) = 0.72$, 0.81 and 0.74 for quantum dots 1, 2 and 3, respectively.

To analyse these data, we fitted the function $M(\Delta t) = 0.5[1 - a \exp(-|\Delta t|/\tau_m)]$, where the fitting parameters a and τ_m characterize the depth and the width of the coincidence dip, respectively. The fits, shown as solid lines in Fig. 4, match the data well. For an ideal spontaneous-emission source, with instantaneous initial excitation and no decoherence, a would differ from 1 only because of imperfections in the optical set-up, and τ_m would be equal to the spontaneous emission lifetime. The fitted values of τ_m we obtain (in ps) are 80, 187 and 378 for quantum dots 1, 2 and 3, respectively. These values agree quite well with the spontaneous emission decay lifetimes τ_s obtained in Fig. 2a (see also Table 1). For quantum dots 1 and 3, this result is surprising, given the short coherence lengths τ_c listed above. We conclude that, for quantum dots 1 and 3, the primary spectral broadening mechanism occurs on a timescale much longer than 2 ns. Such a ‘spectral diffusion’ effect could occur owing to charge fluctuations in the vicinity of the quantum dot, for example²².

For quantum dot 2, we calculate a mean two-photon overlap of at least 0.81. The remaining imperfection could arise from several decoherence mechanisms. When the quantum dot is first excited by a laser pulse, the generated electron–hole pair is initially in an excited state, and must relax to its lowest state through phonon emission before a photon can be emitted at the proper wavelength. The ratio of this relaxation time, which could be as long as tens of picoseconds, to the lowest-state radiative lifetime could limit the performance of this source. Decoherence by phonons^{24,25} is another possible mechanism, though we see little temperature dependence from 3 to 7 K. Finally, the spectral diffusion mechanism noted above could also potentially contribute to decoherence on short timescales.

The two-photon interference effect that we observe indicates a large enough degree of photon indistinguishability to perform interesting quantum-optical experiments. The performance of most schemes based on two-photon interference depends on the same wave-packet overlap as measured here. For example, for a single-photon implementation of a scheme to generate single pairs of polarization-entangled photons¹⁷, the polarization correlation would ideally be unity in the horizontal/vertical basis, and 0.81 in the $+45^\circ/-45^\circ$ basis, violating Bell’s inequality. We hope that other applications, such as quantum teleportation and quantum logic gates, will become feasible as the performance of single-photon sources continues to improve. $\square$

Received 27 June; accepted 23 August 2002; doi:10.1038/nature01086.

- De Martini, E., Di Giuseppe, G. & Marrocco, M. Single-mode generation of quantum photon states by excited single molecules in a microcavity trap. *Phys. Rev. Lett.* **76**, 900–903 (1996).
- Brunel, C., Lounis, B., Tamarat, P. & Orrit, M. Triggered source of single photons based on controlled single molecule fluorescence. *Phys. Rev. Lett.* **83**, 2722–2725 (1999).
- Lounis, B. & Moerner, W. E. Single photons on demand from a single molecule at room temperature. *Nature* **407**, 491–493 (2000).

4. Kim, J., Benson, O., Kan, H. & Yamamoto, Y. A single-photon turnstile device. *Nature* **397**, 500–503 (1999).
5. Beveratos, A. *et al.* Room temperature stable single-photon source. *Eur. Phys. J. D* **18**, 191–196 (2002).
6. Kuhn, A., Hennrich, M. & Rempe, G. Deterministic single-photon source for distributed quantum networking. *Phys. Rev. Lett.* **89**, 067901 (2002).
7. Michler, P. *et al.* A quantum dot single-photon turnstile device. *Science* **290**, 2282–2285 (2000).
8. Santori, C., Pelton, M., Solomon, G., Dale, Y. & Yamamoto, Y. Triggered single photons from a quantum dot. *Phys. Rev. Lett.* **86**, 1502–1505 (2001).
9. Zwiller, V. *et al.* Single quantum dots emit single photons at a time: antibunching experiments. *Appl. Phys. Lett.* **78**, 2476–2478 (2001).
10. Yuan, Z. *et al.* Electrically driven single photon source. *Science* **295**, 102–105 (2002).
11. Gérard, J.-M. & Gayral, B. Strong Purcell effect for InAs quantum boxes in three-dimensional solid-state microcavities. *J. Lightwave Technol.* **17**, 2089–2095 (1999).
12. Knill, E., Laflamme, R. & Milburn, G. J. A scheme for efficient quantum computation with linear optics. *Nature* **409**, 46–52 (2001).
13. Fearn, H. & Loudon, R. Theory of two-photon interference. *J. Opt. Soc. Am. B* **6**, 917–927 (1989).
14. Hong, C. K., Ou, Z. Y. & Mandel, L. Measurement of subpicosecond time intervals between two photons by interference. *Phys. Rev. Lett.* **59**, 2044–2046 (1987).
15. Bennett, C. H. & Brassard, G. *Proc. IEEE Int. Conf. on Computers, Systems and Signal Processing* 175–179 (IEEE, New York, 1984).
16. Bouwmeester, D., Ekert, A. & Zeilinger, A. *The Physics of Quantum Information* 49–92 (Springer, Berlin, 2000).
17. Shih, Y. H. & Alley, C. O. New type of Einstein–Podolsky–Rosen–Bohm experiment using pairs of light quanta produced by optical parametric down conversion. *Phys. Rev. Lett.* **61**, 2921–2924 (1988).
18. Moreau, E. *et al.* Single-mode solid-state single photon sources based on isolated quantum dots in pillar microcavities. *Appl. Phys. Lett.* **79**, 2865–2867 (2001).
19. Solomon, G. S., Pelton, M. & Yamamoto, Y. Single-mode spontaneous emission from a single quantum dot in a three-dimensional microcavity. *Phys. Rev. Lett.* **86**, 3903–3906 (2001).
20. Borri, P. *et al.* Ultralong dephasing time in InGaAs quantum dots. *Phys. Rev. Lett.* **87**, 157401 (2001).
21. Gammon, D., Snow, E. S., Shanabrook, B. V., Katzer, D. S. & Park, D. Homogeneous linewidths in the optical spectrum of a single gallium arsenide quantum dot. *Science* **273**, 87–90 (1996).
22. Bayer, M. & Forchel, A. Temperature dependence of the exciton homogeneous linewidth in $\text{In}_{0.60}\text{Ga}_{0.40}\text{As}/\text{GaAs}$ self-assembled quantum dots. *Phys. Rev. B* **65**, 041308(R) (2002).
23. Kulakovskii, V. D. *et al.* Fine structure of biexciton emission in symmetric and asymmetric CdSe/ZnSe single quantum dots. *Phys. Rev. Lett.* **82**, 1780–1783 (1999).
24. Besombes, L., Kheng, K., Marsal, L. & Mariette, H. Acoustic phonon broadening mechanism in single quantum dot emission. *Phys. Rev. B* **63**, 155307 (2001).
25. Fan, X., Takagahara, T., Cunningham, J. E. & Wang, H. Pure dephasing induced by exciton–phonon interactions in narrow GaAs quantum wells. *Solid State Commun.* **108**, 857–861 (1998).

Acknowledgements We thank A. Scherer and T. Yoshie for access to the CAIBE system and for their help in fabrication of the structures; J. Plant for assistance with SEM imaging; and B. C. Sanders for discussions. This work was supported in part by MURI; G.S.S. was supported by DARPA, ARO and JST.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.S. (e-mail: chars@stanford.edu).

Superconductivity in compressed lithium at 20 K

Katsuya Shimizu^{*†}, Hiroto Ishikawa^{*}, Daigoroh Takao^{*}, Takehiko Yagi[‡] & Kiichi Amaya^{*†}

^{*} Department of Physical Science, Graduate School of Engineering Science; and [†] Research Center for Materials Science at Extreme Conditions, Osaka University, Toyonaka, Osaka 560-8531, Japan

[‡] The Institute for Solid State Physics, University of Tokyo, Kashiwa-shi, Chiba 277-8581, Japan

Superconductivity at high temperatures is expected in elements with low atomic numbers, based in part on conventional BCS (Bardeen–Cooper–Schrieffer) theory¹. For example, it has been predicted that when hydrogen is compressed to its dense metallic phase (at pressures exceeding 400 GPa), it will become superconducting with a transition temperature above room tempera-

ture². Such pressures are difficult to produce in a laboratory setting, so the predictions are not easily confirmed. Under normal conditions lithium is the lightest metal of all the elements, and may become superconducting at lower pressures^{3,4}; a tentative observation of a superconducting transition in Li has been previously reported⁵. Here we show that Li becomes superconducting at pressures greater than 30 GPa, with a pressure-dependent transition temperature (T_c) of 20 K at 48 GPa. This is the highest observed T_c of any element; it confirms the expectation that elements with low atomic numbers will have high transition temperatures, and suggests that metallic hydrogen will have a very high T_c . Our results confirm that the earlier tentative claim⁵ of superconductivity in Li was correct.

Previous theory⁶ has predicted that dense Li will undergo a new structural transition towards a ‘paired-atom’ phase at pressures near

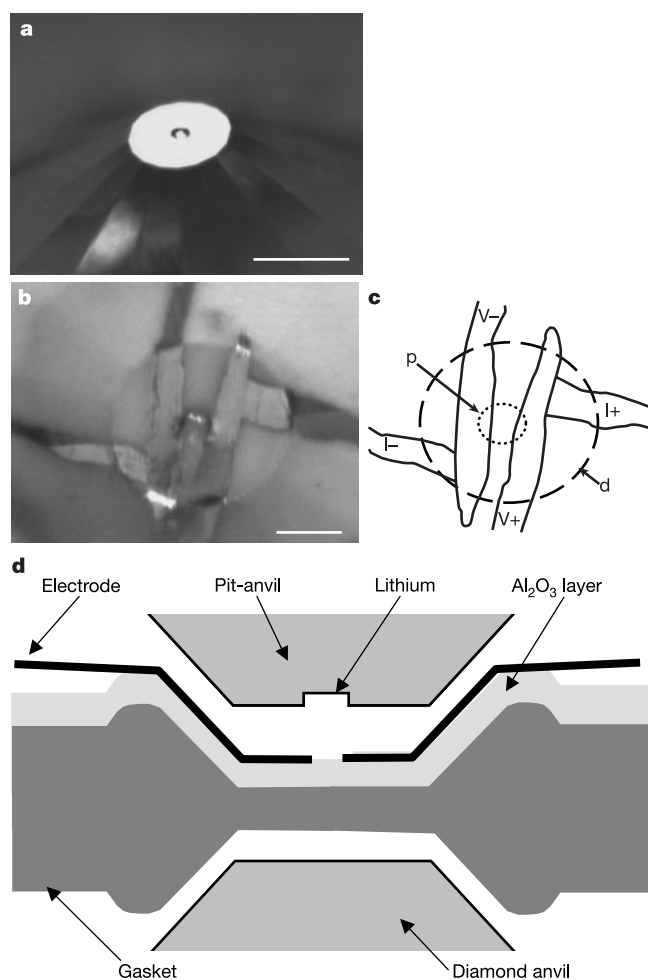


Figure 1 Arrangement of sample and electrodes on the diamond anvils. **a**, Photograph of a pit, 50 μm in diameter and 7 μm deep, on the 300- μm pressure surface of the synthetic type Ib diamond anvil. The pit was prepared by a focused ultraviolet beam from pulsed KrF-excimer laser, wavelength 248 nm. Scale bar, 0.3 mm. **b**, Electrodes on a thin aluminium oxide layer. Two platinum-film electrodes, 5- μm thick, are placed to touch the sample in the pit; scale bar, 0.1 mm. **c**, Schematic drawing of **b**. A quasi-four-wire electrical resistance measurement was performed. The measured resistivity included one of the platinum films which were placed in series connecting to the sample in the pit (p) on the diamond-anvil surface (d). A direct current of 1 mA is applied through I+ to I–, and the voltage drop between V+ and V– is recorded. **d**, Schematic drawing of the cross-section of our set-up at the top of the diamonds anvils. Ruby chips are located the bottom of the pit; the pressure was controlled by helium gas, and determined by a conventional ruby-fluorescence method through the optical windows of the cryostat.

100 GPa, and X-ray diffraction measurements⁷ have shown that new structural changes occur above 40 GPa. Recent theoretical calculations⁸ for superconductivity in Li predicted that T_c may reach a maximum of 80 K in the *c*/16 phase. According to structural calculations⁷, this phase is stable up to 165 GPa, and then transforms to a paired structure. However, in comparison with other alkali metals, Li has fewer experimental reports on its high-pressure transport properties. This could be because of its high chemical activity, and its reaction with the diamond used in high-pressure diamond-anvil cells (DACs; refs 7, 9). We also had tried a number of such experiments, and failed to maintain sample purity at pressures above 30 GPa; contamination was caused by chemical reaction, and leaks occurred through the spaces between the gasket and the diamond anvil. To prevent leaking of the Li sample, we used a diamond-anvil with a 'pit'; the experimental set-up is shown in Fig. 1. The sample was cut into a suitable size from high-purity ribbon (⁷Li, 99.999%, Jonson-Matthey), and loaded into a DAC. All of these treatments were done in an argon atmosphere in a water- and oxygen-free glove-box. We applied pressure not exceeding 10 GPa in the glove-box, and then the pressure-cell was connected to a cryostat. Further pressure was applied at temperatures lower than 50 K in order to minimize the chemical reactions between the

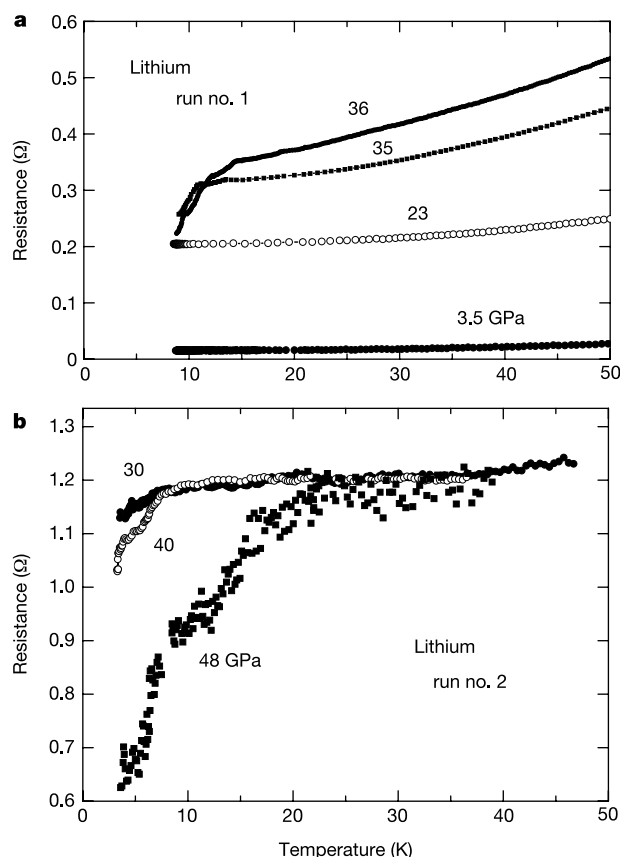


Figure 2 Temperature dependence of the electrical resistance of lithium measured at temperatures below 50 K. We obtained data from different runs, numbers 1 & 2. **a**, The pressure was increased at around 50 K, and the same curve was obtained in the cooling and warming directions in this run. A resistance drop at 13 K of about 30% was observed at 35 GPa. After increasing pressure up to 36 GPa at 50 K, the residual resistance increased and the onset of the drop slightly shifted up to 14 K with a larger transition step. During further compression from 36 GPa, the diamonds in the DAC broke. **b**, In the case of run no. 2, the pressure was changed at 4 K and the increment of the resistance during compression was negligible. The observed onset temperature of the drop shifted to higher temperature with increasing pressure, and reached to 20 K at 48 GPa.

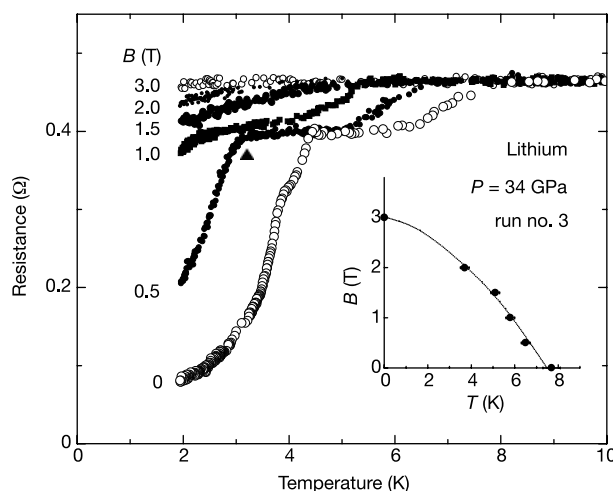


Figure 3 The effect of magnetic fields on the behaviour of the resistance drop in run no. 3. The transition was suppressed by application of a magnetic field B , and disappeared above 3 T, which is strong evidence for superconductivity. We regard the remaining resistance as the contact resistance between the sample and electrode, including the un-superconducting part of the sample in series with the measuring path. The onset temperature of 7.2 K at 0 T was close to the value for lead (Pb) at atmospheric pressure: but we used no Pb in the measuring path, and the critical magnetic field was too high for Pb, as shown in the inset. One plateau was clearly observed up to 1.0 T (marked by a triangle for 0.5 T). The plateau was also suppressed by the magnetic field. We observed the plateau throughout run numbers 1–4; however, we have no explanation for its origin except for a pressure gradient in the sample.

Li and the anvil. Some part of the sample spread out from the pit at the beginning of compression, but no further leakage was observed; the pit confined the Li in the anvil. The insulated layer succeeded in keeping electrical insulation between the sample and the gasket in all of the present runs.

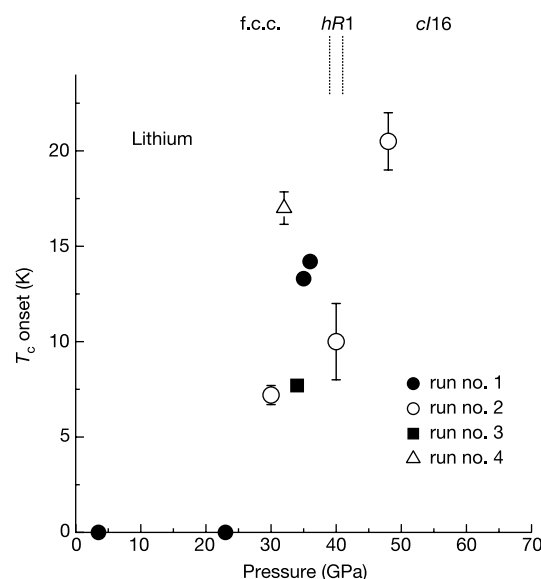


Figure 4 Pressure dependence of the superconducting onset temperature, T_c onset, for Li. Dotted lines at the top of the figure show two structural phase boundaries at 200 K determined by a low-temperature X-ray study⁷: face-centred cubic (f.c.c.) to *h*R1, and *h*R1 to *c*/16, at 39 and 41 GPa, respectively. It is uncertain whether the phase boundary at 200 K meets the one at low temperature where the superconductivity was observed.

Pressure was controlled at around 50 K, and curves of resistance versus temperature were obtained at fixed pressures (3.5–48 GPa; Fig. 2). At pressures of 3.5 and 23 GPa, the measured resistance showed a normal metallic behaviour. The residual resistance increased with compression, which could be explained by thinning of the sample and possible chemical reactions at the contact point between the sample and the electrodes. At a pressure of 35 GPa, a drop of resistance was observed at around 13 K, and the onset temperature of the drop shifted slightly to higher temperature at 36 GPa. At 48 GPa (run no. 2), the width of the resistance drop increased and the onset temperature reached 20 K (Fig. 2a). The loss of resistance (50%) suggested that the drop was caused by the superconducting transition. Figure 3 shows the magnetic field dependence of the resistance drop at 34 GPa (run no. 3), which is strong evidence for the superconducting transition; the drop shifted to lower temperature when the magnetic field was applied, and was fully suppressed above 3 T. Observation of the Meissner signal is indispensable for proving the existence of superconductivity, but we had experimental difficulties in performing magnetic measurements with the gas-controlled DAC, and have not obtained this signal. After releasing the pressure, the sample showed its original reflection of light, which indicated that chemical reaction was negligible during the present experiments. We conclude that the drop of resistance is due to the onset of superconductivity of Li.

Figure 4 shows the pressure dependence of the superconducting onset temperature of Li observed in four different runs (nos 1–4). Data are scattered, but the onset temperature clearly shows (on average) a rise with pressure. Lin and Dunn⁵ had previously reported high-pressure and low-temperature electrical resistance measurements on Li, and observed a drop of resistance at around 7 K in a similar pressure region to our results. Although they concluded that this drop of resistance was caused by a phase transition, which could include (as they tentatively proposed) the onset of superconductivity, here we suggest the resistance drop observed by these workers was indeed likely to have been from superconductivity. The present result establishes that compressed Li is a superconductor.

According to theoretical predictions^{7,8} we may be able to achieve a T_c higher than 20 K on further compression. High-pressure investigations of Li, the first superconducting element of the periodic table, should help the understanding both of the fundamental properties of metals, and of the possible room-temperature superconductivity in metallic hydrogen². □

Received 11 June; accepted 26 August 2002; doi:10.1038/nature01098.

1. Bardeen, J., Cooper, L. N. & Schrieffer, J. R. Theory of superconductivity. *Phys. Rev.* **108**, 1175–1204 (1957).
2. Richardson, C. F. & Ashcroft, N. W. High temperature superconductivity in metallic hydrogen: Electron-electron enhancements. *Phys. Rev. Lett.* **78**, 118–121 (1997).
3. Allen, P. B. & Cohen, M. L. Pseudopotential calculation of the mass enhancement and superconducting transition temperature of simple metals. *Phys. Rev.* **187**, 525–538 (1969).
4. Richardson, C. F. & Ashcroft, N. W. Effective electron-electron interactions and the theory of superconductivity. *Phys. Rev. B* **55**, 15130–15145 (1997).
5. Lin, T. H. & Dunn, K. J. High-pressure and low-temperature study of electrical resistance of lithium. *Phys. Rev. B* **33**, 807–811 (1986).
6. Neaton, J. B. & Ashcroft, N. W. Pairing in dense lithium. *Nature* **400**, 141–144 (1999).
7. Hanfland, M., Syassen, K., Christensen, N. E. & Novikov, D. L. New high-pressure phase of lithium. *Nature* **408**, 174–178 (2000).
8. Christensen, N. E. & Novikov, D. L. Predicted superconductive properties of lithium under pressure. *Phys. Rev. Lett.* **86**, 1861–1864 (2001).
9. Hanfland, M., Loa, I., Syassen, K., Schwarz, U. & Takemura, K. Equation of state of lithium to 21 GPa. *Solid State Commun.* **112**, 123–127 (1999).

Acknowledgements We thank H. Goto for technical support in machining the pit-anvils. This work was supported by a Grant-in-Aid for COE (Center of Excellence) Research and Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to K.S. (e-mail: kshimizu@mp.es.osaka-u.ac.jp).

Propagation of the polyamorphic transition of ice and the liquid–liquid critical point

Osamu Mishima & Yoshiharu Suzuki

Advanced Materials Laboratory (AML), National Institute for Materials Science (NIMS), 1-1 Namiki, Tsukuba, 305-0044 Japan

Water has a rich metastable phase behaviour that includes transitions between high- and low-density amorphous ices, and between high- and low-density supercooled liquids. Because the transitions occur under conditions where crystalline ice is the stable phase, they are challenging to probe directly. In the case of the liquids, it remains unclear¹ whether their mutual transformation at low temperatures is continuous^{2,3}, or discontinuous^{4,5} and terminating at a postulated second critical point of water that is metastable with respect to crystallization. The amorphous ices are more amenable to experiments^{6–8}, which have shown that their mutual transformation is sharp and reversible. But the non-equilibrium conditions of these studies make a firm thermodynamic interpretation of the results difficult. Here we use Raman spectroscopy and visual inspection to show that the transformation of high-density to low-density amorphous ices involves the propagation of a phase boundary—a region containing a mixture of both ices. We find that the boundary region becomes narrower as the transformation progresses, and at higher transformation temperatures. These findings strongly suggest that the polyamorphic ice transition is discontinuous; a continuous transformation should occur uniformly over the entire sample⁹. Because the amorphous ices are structurally similar to their supercooled liquid counterparts, our results also imply that the liquids transform discontinuously at low temperatures and thus support the liquid–liquid critical-point theory^{4,5}.

We used pressure-induced amorphization of deionized H₂O ice at ~1 GPa, 77 K, to obtain high-density amorphous ice (HDA). The sample was then heated at 1.5 GPa to 150 K at ~3 K min^{−1} to make a large homogeneous HDA block; this was then ‘recovered’ at 100 kPa, 77 K (refs 10–12), yielding ‘as-recovered’ HDA as a transparent solid with few cracks.

A lump of the ‘as-recovered’ HDA, 1–2 mm in size, was placed at 77 K in a hole in an indium block attached to a cryostat, and heated in vacuum at 0.5 K min^{−1} to the annealing temperature T_{anneal} . The sample was kept at this temperature for about 3–3.5 h, and then cooled to 25 K; throughout this procedure, microscopic Raman spectra were taken to identify the ice phases¹³ of the sample. The heating–annealing–cooling sequence was then repeated using a higher T_{anneal} .

Annealing (ageing) at T_{anneal} releases excessive stress: the HDA structure changes gradually and irreversibly, until the changes gradually cease with time¹⁴ as evident from the Raman spectra and a decrease in the heat evolved from HDA per unit time (not shown). The Raman peak frequency of the annealed HDA sample changes with temperature almost reversibly between T_{anneal} and 25 K at a rate similar to that of low-density amorphous ice (LDA) and ice Ih (ref. 15), indicating that its structure is stable below T_{anneal} (ref. 14).

The Raman spectra of the sample at 25 K after the 3–3.5 h annealing at different T_{anneal} are shown in Fig. 1, and the corresponding peak frequencies, relative peak intensities, and estimated specific volumes at T_{anneal} are shown in Fig. 2. The HDA structure relaxes on annealing at ~90–115 K, and HDA transforms to LDA at 115 ± ~0.5 K, in agreement with previous results^{12,16–18}. Although the transformation from HDA into LDA is associated with heat

release¹⁶, we believe that this does not lead to any significant heating of the sample itself because the transformation proceeds slowly.

Although the peak frequency and the specific volume of HDA change increasingly as T_{anneal} approaches 115 K (Fig. 2), the Raman spectra exhibit a common, broad profile (Fig. 1). The profiles measured at several different points on the surface of a given relaxed HDA sample are always identical within experimental error; any stresses remaining must therefore be homogeneously distributed throughout the HDA sample. Annealing to temperatures less than 115 K is accordingly accompanied by uniform sample swelling that causes no cracks; annealing at 115 K, in contrast, results in heterogeneous swelling and cracking, belying a monotonous change¹⁴ to LDA.

The 25-K spectra of HDA annealed near 112 K (Fig. 1) are, within experimental error, identical to the 25-K spectra^{13,19} of HDA samples that are not initially heat-treated at 1.5 GPa and also not annealed, indicating that these samples have similar structures, even though they differ subtly¹². As annealing of HDA at 112 K at 100 kPa and heating HDA to 150 K at 1.5 GPa induces a gradual change to another HDA, their structures appear to change mutually and continuously. This suggests that different HDA states lie within the same HDA megabasin in configuration space¹⁰.

The temporal evolution of the spatial distribution of the different ice states during annealing near 115 K is illustrated in Fig. 3. The HDA-to-LDA transformation starts in the left part of the ice sample and propagates to the right, accompanied by an increase in sample volume from left to right. The Raman spectra of the left part of the ice sample shown in Fig. 3b are similar to the Raman spectrum of LDA, whereas spectra taken of the right part of the sample resemble the HDA spectrum. The region between the LDA and HDA phases shows continuously changing intermediate spectra, which can be expressed as a linear combination of the spectra of HDA and LDA (Fig. 4). We classify these ice states by the fraction of the HDA component, α (see Fig. 4 legend for a full definition) and show in the right-hand panels of Fig. 3 how α changes along a line in the transformation direction (indicated by the X-X' arrow in the left panels). The slope of α in the intermediate-state region (the slope of the broken lines in the right panels in Fig. 3) becomes steeper with time and, hence, the boundary region separating HDA from LDA narrows gradually. This suggests that the intermediate state changes more quickly to LDA than HDA changes to the intermediate state—a phase-separating behaviour. Large cracks, indicative of severe

stress, appear mainly in the direction perpendicular to the transformation direction (Fig. 3). The high density of cracks in the right part of the ice sample correspond with a particularly steep slope in α observed at that location, suggesting that crack density correlates with the broadness of the boundary region.

In a separate experiment, we partly transformed a sample at 115 K and then kept it at a lower temperature of 111 K and monitored the changes of α along a line in the transformation direction. As illustrated in Fig. 5, the values of α in the left part of the sample, left of point P, decrease over time as the corresponding intermediate states transform into LDA. In contrast, HDA states to the right of this point remain unchanged. This behaviour clearly illustrates that the partially transformed sample segregates into HDA and LDA regions separated by a crack at P. Upon re-heating, the remaining HDA starts to change to the intermediate state (Fig. 5e). The data also show that the phase transformation of HDA to LDA, once initiated at 115 K, triggers complete transformation of the affected sample part, even at 111 K. This 'over-heating' effect is comparable to the 'over-depressurization' of the decompression-induced HDA-to-LDA transformation around 135 K, where the phase transformation, once it has begun, continues at a higher pressure (the transition b in Fig. 2 of ref. 8). The behaviour is consistent with the observed 'coexistence' of HDA and LDA during the phase transformation^{12,20}, which must therefore be discontinuous.

We used higher transformation temperatures to study the effect of temperature on the broadness of the boundary region. After annealing at 108–110 K for ~20 h and confirmation via Raman spectra that the HDA sample was in a state just before the transformation, we rapidly (~10 K min⁻¹) heated the sample and then kept it at a constant temperature between 115 K and 131 K. The transformation to LDA, once it starts, propagates more quickly at higher temperatures, leaving dense major and fine cracks. The transformation completes in ~5 h at

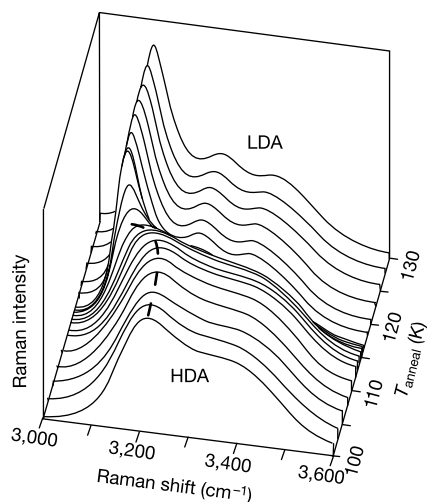


Figure 1 Changes in the Raman spectrum of a high-density amorphous ice sample during annealing at different temperatures T_{anneal} . Total intensity between 2,800 and 3,650 cm⁻¹ in each spectrum is normalized. We measure the spectra to ± 2 cm⁻¹ at 25 K using an argon-ion laser (wavelength 488.0 nm); the laser power in a spot (~20 μ m diameter) on the sample is ~10 mW. Sample temperature is controlled to ± 0.1 K. HDA and LDA, high-density and low-density ice, respectively.

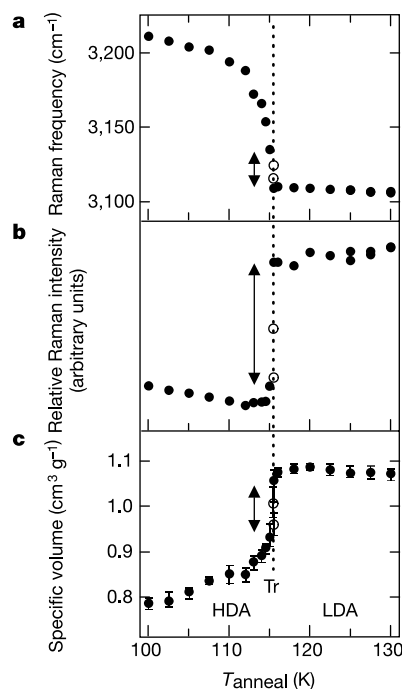


Figure 2 Changes in properties of Raman spectra induced by annealing of amorphous ice at different temperatures. **a**, Raman peak frequency; **b**, relative Raman peak intensity; and **c**, specific volume. Discontinuity at the transformation (Tr) is shown by gaps (arrows). Open circles correspond to the intermediate states at Tr. The specific volume is obtained by observing the sample surface with video equipment. We measure the length between several fixed points on the surface, raise it to the third power to obtain temporary volume, and estimate the specific volume to ~2% by calibrating the volume above 150 K with the specific volume of ice Ih or ice Ic.

115 K and in ~ 0.03 s at 131 K. Raman spectra of a sample transforming at 123 K yield a steep slope of α in an apparently connected part of the sample that soon cracks; the slope is comparable to the steep slope measured for the sample of Fig. 3d.

Because of the speed of the transformations at higher temperatures, Raman measurements are not feasible, and we therefore used video equipment to observe the crack formation (or the volume evolution) during the transformations. Above ~ 132 K the transformation occurs within a single video frame (0.03 s)—making it too fast to observe the propagation. Owing to the variety of transformation directions, sudden crack formation and lack of *in situ* identification of the intermediate states, we have to repeat our observations dozens of times to obtain a meaningful result. We find that the boundary region between the cracked transformed part of the sample and the uncracked and untransformed part of the sample appears more distinct at higher temperatures (Fig. 6), suggesting that the boundary region is narrower, or similar, in width to the boundary region appearing in low-temperature transformations. This is further supported by the appearance of dense

major cracks, which suggest narrow boundary regions (Fig. 3). The propagation and the probable narrowing of the boundary region indicate the relative stability of HDA even at high temperatures.

The presence of a boundary region in partly transformed samples argues for the HDA–LDA transformation having a first-order-like nature⁹, given that continuous structural relaxation of HDA only at the boundary to the LDA region seems very unlikely. In fact, we observe that HDA relaxation always occurs uniformly and essentially without causing cracking, even in cases where quick and large volume expansion accompanies the relaxation. For example, when we heat ‘as-recovered’ 77-K HDA at ~ 20 K min⁻¹, it swells uniformly by 3–5 vol.% within a few seconds just before the transformation to LDA starts to propagate through the sample. The rate of this initial volume expansion is much higher than the rate of the volume expansion accompanying the HDA–LDA transformation at 115 K, yet the relaxation occurs uniformly and without causing cracks. Stresses heterogeneously distributed in HDA must therefore be able to quickly relax and spread uniformly through the sample. Moreover, if the boundary region appearing during the HDA–LDA transformation were caused by relaxation, it should broaden at higher temperatures because structural relaxation processes are usually more uniform at higher temperatures; in contrast and as discussed above, we see evidence for narrowing of the boundary

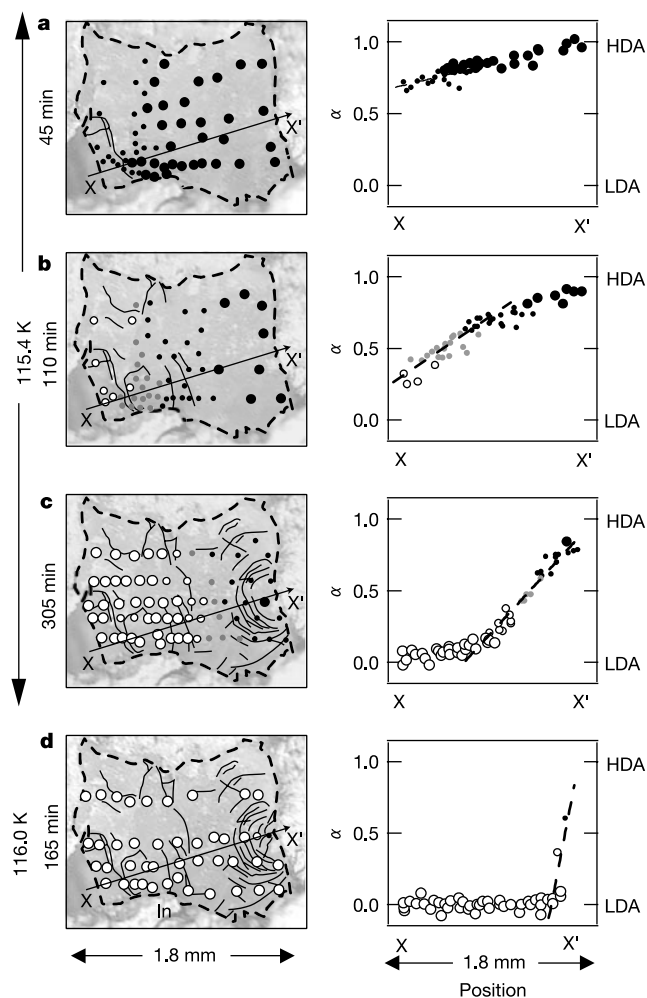


Figure 3 The transformation of HDA to LDA as a function of time. Time given on the left is that elapsed after annealing has started at 115.4 K (a–c) or 116 K (d). We suspend annealing by cooling the ice sample to 25 K, and then measure the Raman spectra at the points indicated on the sample (inside the broken lines) in an indium holder (In). Thin lines, observed major cracks. The surface state at each point is classified, and indicated by the appropriate dot. $\alpha > 0.8$, large black (HDA-like); $0.8 > \alpha > 0.6$, small black; $0.6 > \alpha > 0.4$, grey; $0.4 > \alpha > 0.2$, small white; $0.2 > \alpha$, large white (LDA-like), allowing for experimental uncertainties of α , $\pm \sim 0.1$. Arrow X–X', direction of transformation. Graphs (right) show the progress of α along the line parallel to the arrow. Intermediate states (as mentioned in Fig. 4) are defined as those with $\alpha = 0$ –1. Broken lines, slopes of α .

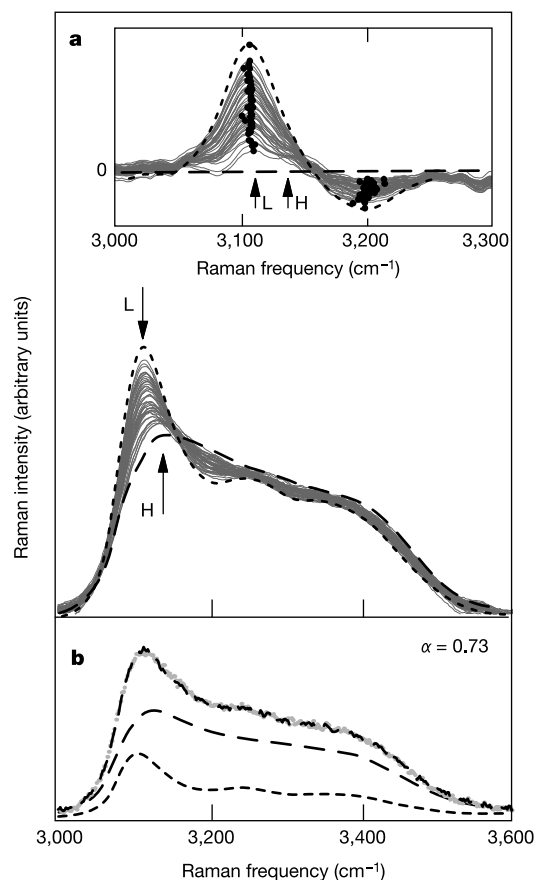


Figure 4 Raman spectra of an intermediate state at 25 K. **a**, All the spectra collected on the sample pictured in Fig. 3b. Inset, the difference in intensity between each spectrum and the HDA spectrum (just before the transformation starts to propagate). Note the isosbestic points and the same peak frequencies in the inset (black points). The arrows, L ($3,112 \pm 2$ cm⁻¹) and H ($3,138 \pm 2$ cm⁻¹), are, respectively, the peak frequencies of the LDA and HDA spectra (the dotted and dashed lines in the main part of **a**), whose difference corresponds to the gap in Fig. 2a. **b**, An example of the linear HDA/LDA combination. We calculate $\alpha/I_{\text{HDA}} + (1 - \alpha)/I_{\text{LDA}}$, where I_{HDA} and I_{LDA} are the HDA and LDA spectra (the dashed and dotted lines), respectively. The result of the fit to experimental data (broken line) agrees with the experimental data (grey dots).

region with increasing temperature.

If the HDA–LDA transformation is first-order-like and therefore involves nucleation and growth of the new phase, the broadness of the low-temperature boundary region and its apparent narrowing at high temperatures are readily explained. That is, HDA will be less viscous at higher temperatures and LDA can thus grow more rapidly in the HDA region, resulting in a relatively narrow HDA–LDA boundary. The boundary region as a manifestation of HDA–LDA phase separation in the ice sample is also consistent with its disappearance in the 111-K experiment (Fig. 5). Moreover, this interpretation readily explains why many cracks appear in ice samples during their transformation, whereas densified silica glasses exhibit monotonic annealing behaviour without cracking (N. Kitamura, personal communication).

The HDA–LDA transformation has also been interpreted as spinodal decomposition, motivated by model calculations^{21,22} that suggest that HDA around 120 K and 100 kPa might be unstable (as opposed to metastable) with respect to LDA. But because spinodal decompositions tend to be system-spanning²³, the observed heterogeneous propagation of the HDA–LDA transformation suggests that the theoretical calculations are not accurate; however, even though never observed so far, it might be possible that a low-

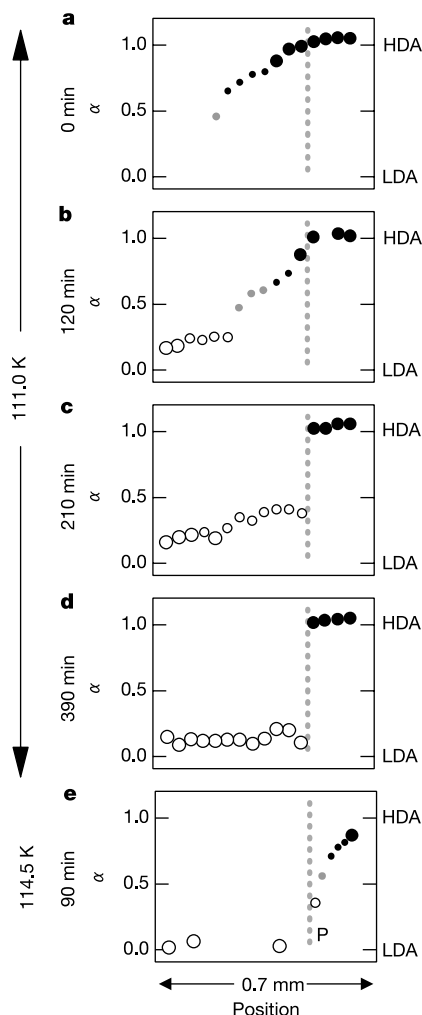


Figure 5 Phase changes in an amorphous sample at 111 K. The HDA–LDA transformation was initiated at 115 K and the sample then cooled to 111 K. Panels **a–e** show the α values along a line parallel to the transformation direction, analogous to the data shown in the right-hand panels in Fig. 3. The partly transformed sample is annealed for the indicated total period at 111 K and cooled for the Raman measurement at 25 K. The HDA–LDA transformation starts again when the sample is re-heated to 114.5 K. A crack grows with time at point P.

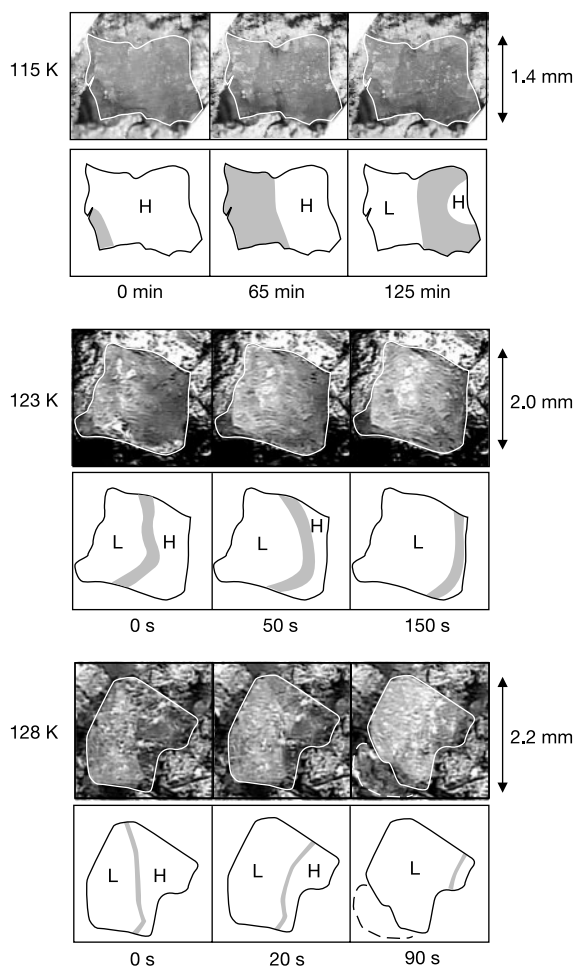


Figure 6 Visual observations of the transformations at 115, 123 and 128 K. Transformations proceed from left to right; times since taking the leftmost image are shown. The boundary region (the grey zone) in the 115-K images is identified using Raman measurements (Fig. 3). The location and width of the boundary regions in the samples transforming at 123 K and 128 K are visually inferred from the boundary between the densely cracked part (white area) and the uncracked part (black area), and are thus unlikely to correspond exactly to the shape given in the schematic figures. The samples just before and just after the 123-K and 128-K transformations are identified by Raman measurements as HDA (H) and LDA (L), respectively. The 123-K sample shown transformed several times quicker than the usual samples at this temperature. The broken line shows the ice separated by a crack. The temperature (or time) at which HDA starts to change triggers the whole transformation, and varies between experiments; however, it can become roughly constant; for example, it occurs at 115 ± 0.5 K in well-annealed HDA in a roughly one-hour timescale.

temperature spinodal decomposition process lacking system-spanning characteristics occurs. At low temperatures and under out-of-equilibrium conditions, it is experimentally very difficult to differentiate clearly and unambiguously between the spinodal decomposition of an unstable state, and the kinetic transition of a metastable state. The detailed mechanism of the HDA–LDA first-order-like transformation thus remains uncertain.

The amorphous phases HDA and LDA easily crystallize on heating. It is thus practically very difficult to observe at temperatures above their expected glass-transition temperatures^{10,19} the equilibrium first-order transformation anticipated to occur around ~ 0.2 GPa on the co-existence line of the low- and high-density liquid waters. But as these equilibrium conditions are approached, the phase separation between HDA and LDA should become more distinct. In fact, this is seen in experiments around ~ 135 K and ~ 0.2 GPa, which reveal LDA-to-HDA ‘over-pressurization’, HDA-

to-LDA 'over-depressurization' and a narrow pressure hysteresis in the reversible transformation of the two phases^{7,8}.

Taken together, our findings strongly argue for the HDA-LDA transformation being a first-order phase transformation. Structurally, HDA and LDA resemble^{24–27} high-density liquid water (HDL) and low-density liquid water (LDL), respectively, and the present results thus also support the liquid-liquid critical-point theory, which holds that HDL and LDL transform into each other through a discontinuous, first-order process. Finally, we note that a clear polyamorphic phase separation has already been observed in the Al_2O_3 - Y_2O_3 system^{28,29}, and may be seen in other network-forming amorphous materials^{30,31}. □

Received 5 July; accepted 3 September 2002; doi:10.1038/nature01106.

1. Debenedetti, P. G. One substance, two liquids? *Nature* **392**, 127–129 (1998).
2. Sastry, S., Debenedetti, P. G., Sciortino, F. & Stanley, H. E. Singularity-free interpretation of the thermodynamics of supercooled water. *Phys. Rev. E* **53**, 6144–6154 (1996).
3. Rebelo, L. P. N., Debenedetti, P. G. & Sastry, S. Singularity-free interpretation of the thermodynamics of supercooled water. II. Thermal and volumetric behavior. *J. Chem. Phys.* **109**, 626–633 (1998).
4. Poole, P. H., Sciortino, F., Essmann, U. & Stanley, H. E. Phase behaviour of metastable water. *Nature* **360**, 324–328 (1992).
5. Franzese, G. & Stanley, H. E. Liquid-liquid critical point in a Hamiltonian model for water: analytic solution. *J. Phys. Condens. Matter* **14**, 2201–2209 (2002).
6. Floriano, M. A., Handa, Y. P., Klug, D. D. & Whalley, E. Nature of the transformations of ice I and low-density amorphous ice to high-density amorphous ice. *J. Chem. Phys.* **91**, 7187–7192 (1989).
7. Mishima, O., Takemura, K. & Aoki, K. Visual observations of the amorphous-amorphous transition in H_2O under pressure. *Science* **254**, 406–408 (1991).
8. Mishima, O. Reversible first-order transition between two water amorphs at ~ 0.2 GPa and ~ 135 K. *J. Chem. Phys.* **100**, 5910–5912 (1994).
9. Moynihan, C. T. *Ann. NY Acad. Sci.* **484**, 94 (1986); **484**, 287 (1986).
10. Mishima, O. Relationship between melting and amorphization of ice. *Nature* **384**, 546–549 (1996).
11. Suzuki, Y., Takasaki, Y., Tomimaga, Y. & Mishima, O. Low-frequency Raman spectra of amorphous ices. *Chem. Phys. Lett.* **319**, 81–84 (2000).
12. Loerting, T. et al. A second distinct structural "state" of high-density amorphous ice at 77 K and 1 bar. *Phys. Chem. Chem. Phys.* **3**, 5355–5357 (2001).
13. Klug, D. D., Mishima, O. & Whalley, E. Raman spectrum of high-density amorphous ice. *Physica B* **139** & **140**, 475–478 (1986).
14. Tulk, C. A. et al. Structural studies of several distinct metastable forms of amorphous ice. *Science* **297**, 1320–1323 (2002).
15. Sivakumar, T. C., Schuh, D., Sceats, M. G. & Rice, S. A. The 2500–4000 cm^{-1} Raman and infrared spectra of low density amorphous solid water and of polycrystalline ice I. *Chem. Phys. Lett.* **48**, 212–218 (1977).
16. Handa, Y. P., Mishima, O. & Whalley, E. High-density amorphous ice. III. Thermal properties. *J. Chem. Phys.* **84**, 2766–2770 (1986).
17. Klug, D. D., Mishima, O. & Whalley, E. High-density amorphous ice. IV. Raman-spectrum of the uncoupled O-H and O-D oscillators. *J. Chem. Phys.* **86**, 5323–5328 (1987).
18. Gromitskaya, E. L., Stal'gorova, O. V., Brazhkin, V. V. & Lyapun, A. G. Ultrasonic study of the nonequilibrium pressure-temperature diagram of H_2O ice. *Phys. Rev. B* **64**, 094205 (2001).
19. Mishima, O. & Suzuki, Y. Vitrification of emulsified liquid water under pressure. *J. Chem. Phys.* **115**, 4199–4202 (2001).
20. Balagurov, A. M. et al. Neutron-diffraction study of phase-transitions of high-pressure metastable ice-VIII. *JETP Lett.* **53**, 30–34 (1991).
21. Poole, P. H. et al. Effect of hydrogen bonds on the thermodynamic behavior of liquid water. *Phys. Rev. Lett.* **73**, 1632–1635 (1994).
22. Truskett, T. M., Debenedetti, P. G. & Sastry, S. A single-bond approach to orientation-dependent interactions and its implications for liquid water. *J. Chem. Phys.* **111**, 2647–2656 (1999).
23. Debenedetti, P. G. *Metastable Liquids* (Princeton Univ. Press, 1996).
24. Floriano, M. A., Whalley, E., Svensson, E. C. & Sears, V. F. Structure of high-density amorphous ice by neutron diffraction. *Phys. Rev. Lett.* **57**, 3062–3064 (1986).
25. Bellissent-Funel, M.-C. & Bosio, L. A neutron scattering study of liquid D_2O under pressure and at various temperatures. *J. Chem. Phys.* **102**, 3727–3735 (1995).
26. Soper, A. K. & Ricci, M. A. Structures of high-density and low-density water. *Phys. Rev. Lett.* **84**, 2881–2884 (2000).
27. Finney, J. L. et al. Structures of high and low density amorphous ice by neutron diffraction. *Phys. Rev. Lett.* **88**, 225503 (2002).
28. Aasland, S. & McMillan, P. F. Density-driven liquid-liquid phase separation in the system Al_2O_3 - Y_2O_3 . *Nature* **369**, 633–636 (1994).
29. Wilding, M. C. & McMillan, P. F. Polyamorphic transitions in yttria-alumina liquids. *J. Non-Cryst. Solids* **293–295**, 357–365 (2001).
30. Katayama, Y. et al. A first-order liquid-liquid phase transition in phosphorus. *Nature* **403**, 170–173 (2000).
31. Deb, S. K., Wilding, M., Somayazulu, M. & McMillan, P. F. Pressure-induced amorphization and an amorphous-amorphous transition in densified porous silicon. *Nature* **414**, 528–530 (2001).

Acknowledgements We thank S. Sastry and P. G. Debenedetti for discussions and comments, and N. Kitamura for information on silica glass.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for information should be addressed to O.M. (e-mail: mishima.osamu@nims.go.jp).

Directly measured mid-depth circulation in the northeastern North Atlantic Ocean

A. S. Bower*, B. Le Cann†, T. Rossby‡, W. Zenk§, J. Gould||, K. Speer¶
P. L. Richardson*, M. D. Prater‡ & H.-M. Zhang#

* Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

† Laboratoire de Physique des Océans, CNRS, UFR Sciences, BP809, 29285 Brest, France

‡ Graduate School of Oceanography, University of Rhode Island, Kingston, Rhode Island 02881, USA

§ Institut für Meereskunde, Christian-Albrechts-Universität, D-24105 Kiel, Germany

|| Southampton Oceanography Centre, Empress Dock, Southampton SO14 3ZH, UK

¶ Department of Oceanography, Florida State University, Tallahassee, Florida 32306, USA

The circulation of water masses in the northeastern North Atlantic Ocean has a strong influence on global climate owing to the northward transport of warm subtropical water to high latitudes¹. But the ocean circulation at depths below the reach of satellite observations is difficult to measure, and only recently have comprehensive, direct observations of whole ocean basins been possible^{2–4}. Here we present quantitative maps of the absolute velocities at two levels in the northeastern North Atlantic as obtained from acoustically tracked floats. We find that most of the mean flow transported northward by the Gulf Stream system at the thermocline level (about 600 m depth) remains within the subpolar region, and only relatively little enters the Rockall trough or the Nordic seas. Contrary to previous work^{5,6}, our data indicate that warm, saline water from the Mediterranean Sea reaches the high latitudes through a combination of narrow slope currents and mixing processes. At both depths under investigation, currents cross the Mid-Atlantic Ridge preferentially over deep gaps in the ridge, demonstrating that sea-floor topography can constrain even upper-ocean circulation patterns.

Warm, subtropical waters are transported to the northern North Atlantic Ocean, mainly by the Gulf Stream and its poleward extension, the North Atlantic Current (NAC), where they are cooled and transformed into the intermediate and deep water masses that spread throughout the global ocean¹. The pathways and speeds of the currents in this region directly affect the magnitude of this thermohaline circulation, but our ability to accurately measure them has suffered from a lack of widespread direct velocity observations. As part of a large international effort to directly observe the circulation throughout this region using floats², several research groups from the USA, the UK, Germany and France recently collaborated in a major initiative to measure the absolute velocity at two levels in the northeastern North Atlantic using acoustically tracked subsurface floats^{2,7–9} (Fig. 1). Here we present a quantitative view of the mid-depth circulation in this region and highlight some unexpected results.

Figure 2 shows the mean potential temperature (θ) distribution at the two float levels from historical data¹⁰, and sets the hydrographic context for the velocity observations. At the upper level, the warm subtropical thermocline waters spread northward mainly in the northeastern North Atlantic. Two sources for this warm water have been proposed^{5,6,11}: the warm, salty Mediterranean Water in the southeast ($\theta > 10^\circ\text{C}$) that enters the North Atlantic through the Strait of Gibraltar, and the water transported by the Gulf Stream and

Present address: National Climatic Data Center, NOAA, Asheville, North Carolina 28801, USA.

the NAC from the western basin ($5.5^{\circ}\text{C} < \theta < 6.5^{\circ}\text{C}$). Previous work has suggested that some of this warm water enters the Nordic seas (where it is transformed into the cold, dense water masses that fill the deep ocean basins), and that some flows generally anti-clockwise around the subpolar region where it is transformed by wintertime cooling and either leaves southward along the western boundary or recirculates within the subpolar gyre^{5,6,11,12}.

At the lower level, the large-scale temperature contrast from southeast to northwest illustrates the competing influence of warmer, saline Deep Mediterranean Water ($\theta > 7^{\circ}\text{C}$) and cooler, fresher Labrador Sea Water (LSW) ($\theta < 3.4^{\circ}\text{C}$)¹³. LSW spreads into the eastern North Atlantic across the Mid-Atlantic Ridge at about 50°N (refs 13, 14). The influence of warmer, saltier Iceland–Scotland Overflow Water, which enters the northern Iceland basin through the Faroe Bank channel, is apparent along the continental slope south of Iceland and on both flanks of the Reykjanes ridge ($\theta > 3.8^{\circ}\text{C}$)¹⁵.

We used the float velocity data to construct maps of mean streamfunction (Fig. 3). Whereas historical circulation schemes for the upper level in the subpolar North Atlantic relied heavily on assumptions about the deep flow field^{16,17}, Fig. 3a provides directly measured, quantitative information on the mean currents during the sampling period. The NAC flows northward along the western boundary with typical mean speeds of $15\text{--}20\text{ cm s}^{-1}$, to about 52°N . Here it turns sharply clockwise around the ‘Northwest Corner’, and heads eastward across the North Atlantic^{18,19}. The streamlines fan out substantially owing to the branching of the NAC into several weaker currents^{18,20–22}, and the mean speed drops to about 2 cm s^{-1} , but they converge again where the NAC crosses the Mid-Atlantic Ridge, where the average speed is $3\text{--}4\text{ cm s}^{-1}$ (higher over the Charlie-Gibbs fracture zone (CGFZ) at $52\text{--}53^{\circ}\text{N}$, 5 cm s^{-1}). All the streamlines that cross the ridge between 50°N and 53°N turn northward into the Iceland basin, where the typical mean speed drops again to 2 cm s^{-1} . Some streamlines turn sharply

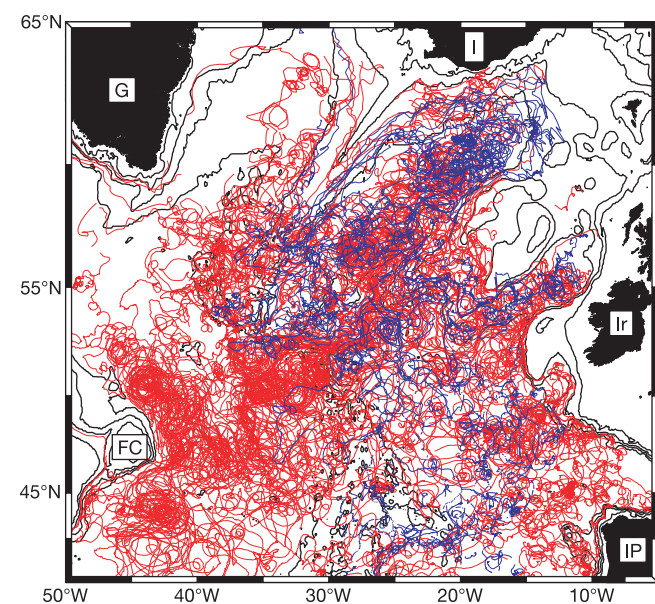


Figure 1 Float trajectories. Shown are trajectories of all 223 acoustically tracked, eddy-resolving, subsurface drifting floats launched in the northern North Atlantic in 1993–95 and 1996–2001, representing 328 float years of data. The floats were deployed at two levels: an upper-ocean thermocline density surface that ranges in depth from $\sim 1,000\text{ m}$ in the subtropics to near the sea surface in the northwestern North Atlantic ($\sigma_{\theta} = 27.5$; red tracks; see also Fig. 2a), and an isobaric layer at $1,500\text{--}1,750\text{ m}$ depth, the core layer of Labrador Sea Water (LSW, blue tracks). Labelled landmarks include Greenland (G), Iceland (I), Ireland (Ir), Iberian peninsula (IP) and FC, Flemish Cap.

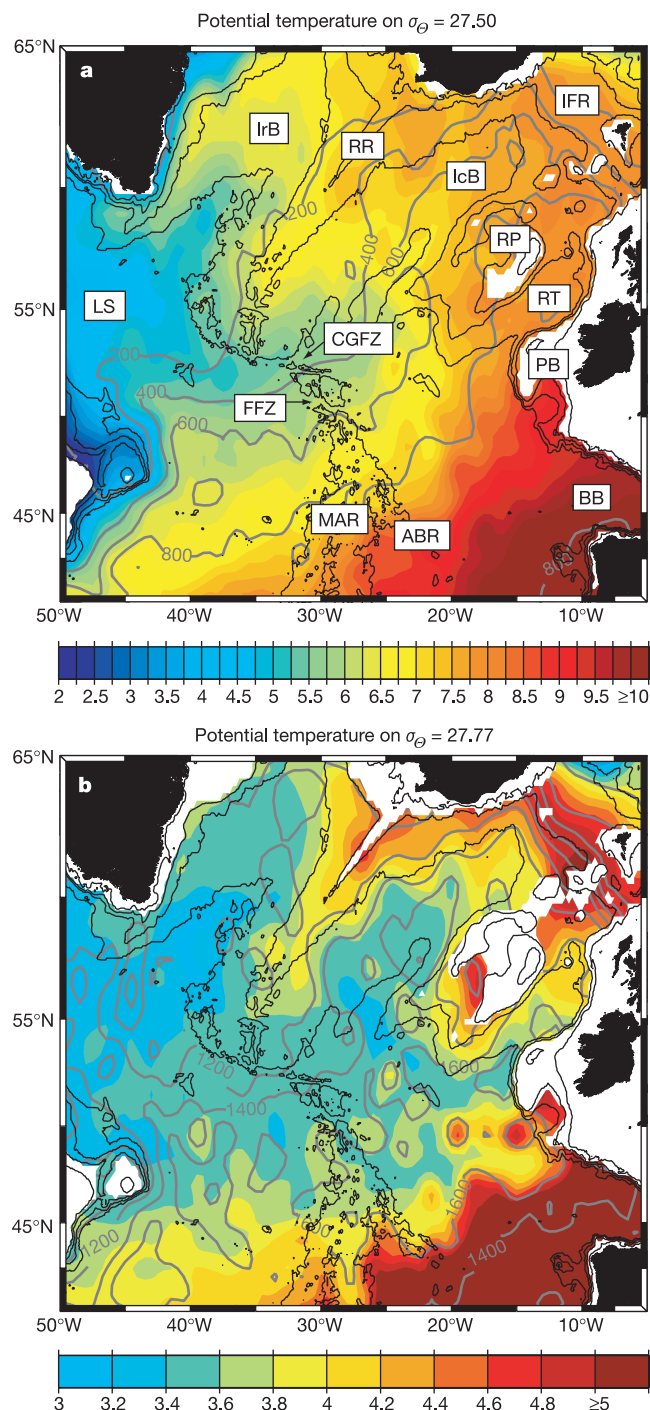


Figure 2 Mean potential temperature (colour shading) on two density surfaces in the subpolar North Atlantic corresponding to the two levels of the float data. **a**, $\sigma_{\theta} = 27.50$, which varies in depth (grey contours) from near the sea surface in the northwestern corner to $\sim 1,000\text{ m}$ near the eastern boundary of the domain. The temperature contour interval is 0.25°C , and bottom depth contours (m) are 200, 1,000, 2,000 and 3,000 (black contours). **b**, Same as **a** but for $\sigma_{\theta} = 27.77$, which ranges in depth from $1,100\text{ m}$ in the northwest to $1,800\text{ m}$ in the southeast, and with temperature contour interval of 0.20°C . Maps are based on historical hydrographic data¹⁰. Temperature is shown on a density surface (rather than constant pressure surface) for the deeper float level to more accurately illustrate water mass distributions. Geographic abbreviations are: Labrador Sea (LS), Irminger basin (IrB), Reykjanes ridge (RR), Iceland basin (IcB), Iceland–Faroes ridge (IFR), Rockall plateau (RP), Rockall trough (RT), Porcupine bank (PB), Bay of Biscay (BB), Azores–Biscay rise (ABR), Mid-Atlantic Ridge (MAR), Faraday fracture zone (FFZ), Charlie-Gibbs fracture zone (CGFZ).

anticlockwise, cross back over the ridge and turn abruptly northward along the western flank of the Reykjanes ridge to form the Irminger Current, while other streamlines penetrate farther north into the Iceland basin and return southwestward along the eastern flank of the ridge before they too cross into the Irminger basin. Speeds along the eastern and western flanks of the ridge are 4–5 and

5–7 cm s^{-1} , respectively. Much weaker mean speeds are indicated in the southeastern quadrant of the study area.

As well as providing a large-scale quantitative view of the thermocline-level circulation, this map reveals a number of unexpected features. First, near the eastern boundary, the mean streamlines are mostly zonal with embedded recirculations, and there is little evidence of a continuous, broad-scale eastern boundary current advecting Mediterranean Water northward from the Iberian peninsula into the subpolar region, as has been described previously^{5,6,23}. Other studies have shown that Mediterranean Water is transported northward along the eastern boundary by a narrow (~50 km) slope current (not resolved here) at least as far north as the northern Bay of Biscay²⁴, but Fig. 3a and individual float tracks indicate that water in this slope current is diverted into the ocean interior and recirculated, particularly at capes and other boundary protrusions. Very few of the floats launched near the eastern boundary south of 52°N drifted northward across that parallel⁷. Mediterranean Water probably penetrates north of 52°N mainly by mixing with the NAC in the complex area southwest of Porcupine bank¹¹. Second, almost all of the NAC mean flow crossing the ridge turns northward into the Iceland basin, and relatively little passes through Rockall trough. (In fact none of the NAC streamlines passes through Rockall trough, but owing to statistical uncertainties in the mean velocity estimates, we cannot rule out the possibility of about $2 \times 10^3 \text{ m}^2 \text{ s}^{-1}$ (transport per unit depth) entering the trough; see Methods). This is consistent with recent hydrographic analysis which suggests that during the low phase of the North Atlantic Oscillation in the late 1990s (characterized by weak westerly winds), the subpolar gyre contracted and the NAC flowed preferentially into the Iceland basin and not the Rockall trough^{25,26}.

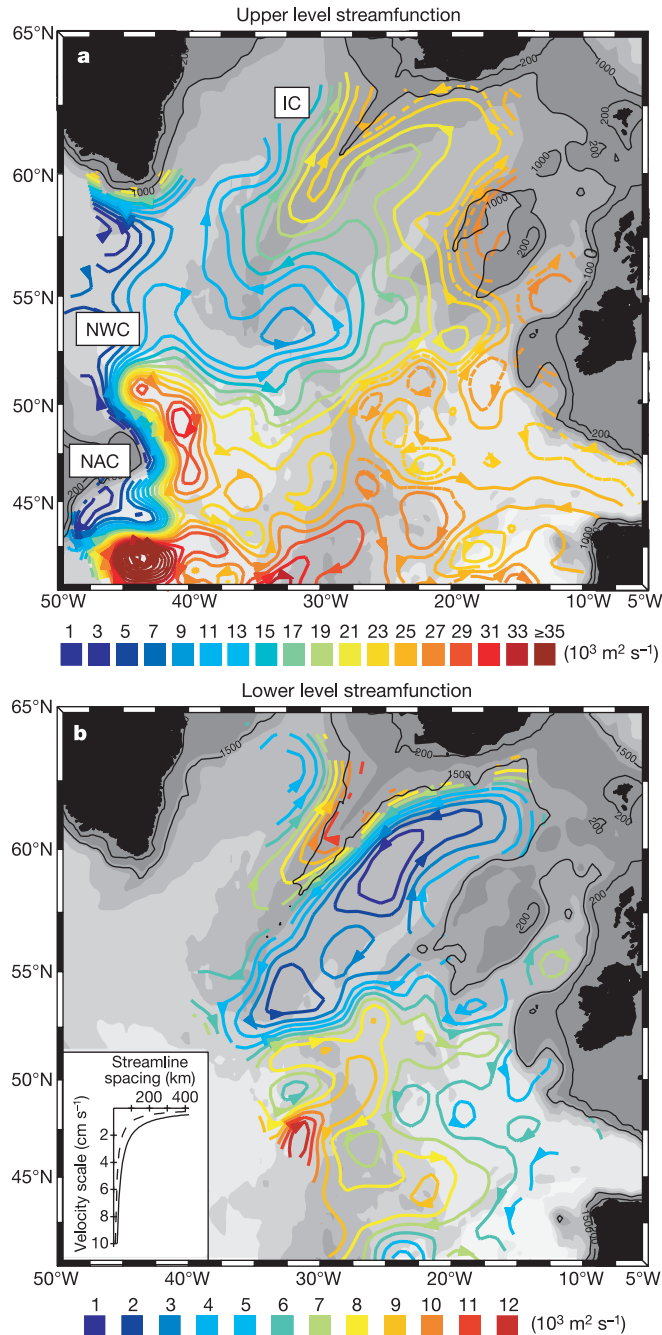


Figure 3 Mean streamfunction for the subpolar North Atlantic from subsurface float data at two levels. **a**, Upper level, thermocline, density $\sigma_\theta = 27.5$; **b**, lower level, Labrador Sea Water, depth 1,500–1,750 m. See Methods for data and mapping details. Arrowheads show the direction of flow along the contours. The inset in **b** gives speed in cm s^{-1} as a function of line separation (solid line for **a**, dashed for **b**). The colour bars below **a** and **b** give volume transport for a one-metre-thick layer. Note that two higher-resolution contours (dashed), 24,000 and 26,000 $\text{m}^2 \text{ s}^{-1}$, have been added near the eastern boundary. Labelled features are: North Atlantic Current (NAC), 'Northwest Corner' (NWC) and Irminger Current (IC).

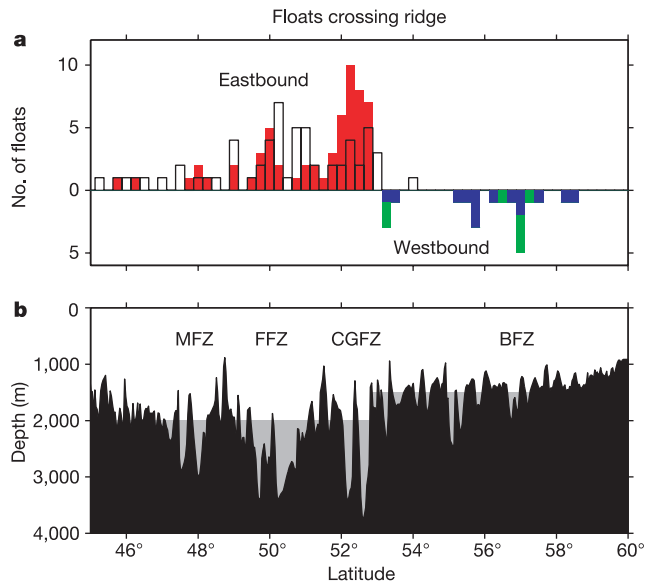


Figure 4 Floats crossing the Mid-Atlantic Ridge. **a**, Distribution of latitudes where floats crossed the Mid-Atlantic Ridge. **b**, Depth of ridge crest as a function of latitude. Floats in the North Atlantic Current at the upper level crossed the ridge eastbound preferentially over the Charlie-Gibbs fracture zone (CGFZ; 31/61), the Faraday fracture zone (FFZ; 11/61), and to a lesser extent, the Maxwell fracture zone (MFZ; 4/61) (red bars; float depths 200–800 m). The distribution of float deployment latitudes, unfilled bars, is significantly different from the distribution of crossing latitudes, suggesting that the floats were funnelled over the various fracture zones. Floats crossed the Reykjanes ridge from east to west north of 53°N at both levels (blue for upper level, green for lower level) mainly through the Bight fracture zone (BFZ) and two unnamed gaps near 53.5°N and 55°N. North of 53°N, the gaps are not aligned east–west and the distribution of float crossing latitudes is therefore somewhat wider around these gaps.

The third unexpected feature is that none of the streamlines, and none of the floats, from the subtropics crosses over the Iceland–Faroes ridge into the Norwegian Sea, even though the $\sigma_\theta = 27.5$ potential density surface is continuous over the ridge (again we cannot preclude some flow over the ridge at this level owing to uncertainties in the mean estimates, but it is probably less than 20% of the NAC flow). This indicates that most of the warm inflow into the Norwegian Sea occurs above the level of these floats¹¹. Most of the flow that enters the Iceland basin from the south must therefore exit into the Irminger basin over the Reykjanes ridge; we find that eight streamlines, or $14 \times 10^3 \text{ m}^2 \text{ s}^{-1}$, cross the ridge first eastward then westward (Fig. 3a). Fourth, we find that the eastward and westward cross-ridge currents flow preferentially over fracture zones (gaps) in the ridge, even at depths up to 1,500 m above the ridge crest. This can be seen in Fig. 3a, but is much clearer in Fig. 4, which shows a histogram of the latitude where upper-level floats crossed the ridge crest (upper panel) relative to the depth of the ridge crest (lower panel). About two-thirds of the eastbound floats (42/61) were ‘funnelled’ over the CGFZ and Faraday fracture zone, and a few over the Maxwell fracture zone. Although it is generally accepted that the northern branch of the NAC is topographically locked to the CGFZ, the southern branches were thought to vary considerably in their crossing latitude^{20–22}. Although fewer upper-level floats crossed the ridge westbound (14), they also preferentially passed over the ridge crest near the deeper gaps, including the Bight fracture zone at 57° N. This could have important implications for the response of the NAC branches to climate-related changes in atmospheric forcing: the shape of the ridge may constrain the current branches to cross within a narrow latitudinal band even for a wide range in forcing, which in turn could affect the circulation and sea surface temperature distribution throughout the subpolar region.

The mean streamfunction map for the lower level (Fig. 3b) illustrates the ‘funnelling’ of LSW from west to east over the CGFZ, with speed $3\text{--}4 \text{ cm s}^{-1}$, some of which appears to be recirculating from the northeast. The eastward current splits into two branches at the Rockall plateau, forming a broad, weak southward flow (typical speed $\sim 0.5 \text{ cm s}^{-1}$) and a somewhat faster northward flow into the Iceland basin ($\sim 2 \text{ cm s}^{-1}$). Higher transport is indicated in the southern branch, consistent with previous hydrographic results¹³. The fastest mean currents at this level, which were found at the northern end of the Iceland basin and along the eastern flank of the Reykjanes ridge, reflect the influence of the Iceland–Scotland overflow. Mean speeds range from 6 to 10 cm s^{-1} along this current. Half of these streamlines cross westward over the Reykjanes ridge and turn northward with speed about 4 cm s^{-1} , while the other half turn eastward at the latitude of the CGFZ. This mean streamfunction map complements a similar map of directly measured LSW circulation at 700 m west of 25° W (ref. 3), and agrees well where they overlap.

Several unanticipated circulation features also emerge from the lower-level map. Two closed, anticlockwise recirculation cells exist adjacent to the boundary current in the western Iceland basin. Although larger in spatial scale, these cells resemble similar recirculations recently observed in the Labrador and Irminger seas³. The individual float trajectories reveal that Iceland–Scotland Overflow Water flows westward into the Irminger Sea not only through the northern part of the CGFZ^{15,27}, but also through deeper gaps in the highly fractured southern one-third of the Reykjanes ridge. Five of the seven deep floats that crossed the ridge westbound did so near the Bight fracture zone at 57° N, while the two others passed westward through a gap just north of the CGFZ.

Similarities in the flow pattern at the two levels include eastward flow over the CGFZ and the clockwise circulation around the Reykjanes ridge. Differences include the vertical structure on opposite sides of the ridge: on the eastern flank, the southwestward flow is significantly faster at the deeper level, while on the western

flank the northeastward flow is faster at the upper level (although the vertical shear is of the same sign on both sides). Also, the closed recirculation in the northwestern Iceland basin at the lower level is absent at the upper level.

The similarity of the streamfunction patterns in Fig. 3 to the details of the sea-floor topography is notable. For example, at the upper level, the NAC closely tracks the 4,000-m isobath in the western basin. At the lower level, the southward flow in the eastern basin is clearly steered clockwise around the Azores–Biscay rise, near 45° N, 20° W (ref. 8). This is expected on the basis of the conservation of potential vorticity, f/H , where f is the Coriolis parameter and H is the fluid layer thickness. Under this conservation constraint, the Mid-Atlantic Ridge represents a significant barrier to inter-basin exchange, and the results shown here suggest that the currents flow preferentially over deeper gaps to cross the ridge. Such topographic steering may limit the ocean’s response to climate variations, locking cross-ridge transports of heat and fresh water to a certain latitude band. We believe that the present data set will provide a standard for numerical simulations of the North Atlantic circulation, and a reference point for future studies of climate variability. □

Methods

Floats

The float data used in this study were collected as part of the US Atlantic Climate Change Experiment (ACCE—a component of the World Ocean Circulation Experiment), the European EUROFLOAT Project, the German Sonderforschungsbereich (SFB) Subpolar Program and the French Actions de Recherche sur la Circulation dans l’Atlantique Nord-Est (ARCANE) during the years 1996–2001. Forty additional floats from the US North Atlantic Current Study (NACS) in 1993–95 were also included¹⁹. The floats were tracked acoustically using an array of 18 moored sound sources that was maintained throughout the North Atlantic by the cooperating groups. Each float recorded at least one position fix and one temperature/pressure observation per day, providing high-resolution trajectories and *in situ* properties.

Float data set

The data set includes both RAFOS floats, which store all their data internally until the end of their missions, and MARVOR floats, which surface every three months to telemeter their stored data and then resubmerge to their operating depths. The upper level data set consists of (1) 116 RAFOS float tracks between 18 and 24 months in duration from the ACCE, (2) 21 MARVOR float tracks up to 60 months long from ARCANE, and (3) 40 RAFOS floats, mostly 10 months in length, from NACS, providing a total of 243 float years of data. The ACCE and NACS floats were made isopycnal by matching their compressibility to that of sea water, and making their coefficient of thermal expansion an order of magnitude less than that of sea water²⁸. These floats were ballasted for the $\sigma_\theta = 27.5$ density surface (neutral surface $\gamma_n \approx 27.6$), which ranges in depth from near the sea surface in the Irminger and Labrador seas, to about 1,000 m adjacent to the European continent (Fig. 2a). Combining these isopycnal floats with the isobaric ARCANE floats, which were ballasted for 1,000 m, is justified because the latter were mainly confined to the southeastern part of the study region, where the 27.5 density surface is nearly 1,000 m deep. The deep floats comprise 25 SFB RAFOS floats at 1,500 m and 21 EUROFLOAT MARVOR floats at 1,750 m for a total of 85 float years. Combining the 1,500- and 1,750-m floats into one layer is justified on the basis of the relative lack of vertical shear between these depths.

Data analysis

The mean velocity at the two levels was estimated by grouping all the float velocity observations according to location into bins 110 km square in flat areas, and somewhat skewed along the isobaths over the continental slope and the Mid-Atlantic Ridge because the mean flows are often aligned along the topography⁴. The data density (N) exceeds 100 float days per bin in much of the North Atlantic in the upper layer, and in the eastern North Atlantic in the lower layer. The resulting mean velocity vectors were mapped into a scalar streamfunction field using standard objective analysis (OA) techniques²⁹. By definition the streamfunction fields produced by the OA are non-divergent: the divergent part is of the order of Ro smaller and thus within the noise (Ro is the Rossby number, U/L , where U and L are typical values for horizontal velocity and length scale). The covariance function has the form $f(r) = [1 + (r/r_o)] \exp(-r/r_o)$, where r is radial distance and r_o is chosen to be 100 km. To help ensure the robustness of the mean streamfunction fields, only areas where the data density exceeds 10 float days per bin were mapped.

Uncertainties

To approximate the uncertainty in the transport per unit depth entering Rockall trough and crossing over the Iceland–Faroes ridge, we first estimate the standard error in the mean velocity. Using $N = 100$ d with an average lagrangian integral timescale (T_I) of five days^{8,30} and a standard deviation (s.d.) around the mean velocity of $\sim 7 \text{ cm s}^{-1}$ (typical for

the regions south of Rockall trough and the Iceland–Faroes ridge), in $\text{s.d.}/(N/T_1)^{1/2}$, the standard error is $\sim 1.5 \text{ cm s}^{-1}$. This translates to an uncertainty in transport per unit depth across the mouth of Rockall trough (two bins) and the Iceland–Faroes ridge (four bins) of $\sqrt{[2, 4] \times (110 \times 10^3 \text{ m}) \times (0.015 \text{ m s}^{-1})} \approx [2, 3] \times 10^3 \text{ m}^2 \text{ s}^{-1}$, where the numbers in brackets refer to the Rockall trough and the Iceland–Faroes ridge, respectively. This would be equivalent to 2.0–2.5 streamlines entering Rockall trough or crossing the Iceland–Faroes ridge, respectively.

Received 10 June; accepted 19 August 2002; doi:10.1038/nature01078.

- Siedler, G., Church, J. & Gould, J. (eds) *Ocean Circulation and Climate—Observing and Modelling the Global Ocean* International Geophysics Series 77 (Academic, San Diego, 2001).
- Davis, R. E. & Zenk, W. in *Ocean Circulation and Climate—Observing and Modelling the Global Ocean* International Geophysics Series 77 (eds Siedler, G., Church, J. & Gould, J.) 123–139 (Academic, San Diego, 2001).
- Lavender, K. L., Davis, R. E. & Owens, W. B. Direct velocity measurements describe a new circulation regime in the Labrador and Irminger Seas. *Nature* **407**, 66–69 (2000).
- Davis, R. Preliminary results from directly measuring mid-depth circulation in the tropical and South Pacific. *J. Geophys. Res.* **103**, 24619–24639 (1998).
- Reid, J. L. On the mid-depth circulation and salinity field in the North Atlantic Ocean. *J. Geophys. Res.* **83**, 5063–5067 (1978).
- Reid, J. L. On the contribution of Mediterranean Sea outflow to the Norwegian–Greenland Sea. *Deep-Sea Res.* **26**, 1199–1223 (1979).
- Bower, A. S. *et al.* Warm-water pathways in the Subpolar North Atlantic: An overview of the ACCE RAFOS float programme. *WOCE Int. Newsl.* **38**, 14–16 (2000).
- Speer, K., Gould, J. & LaCasce, J. Year-long float trajectories in the Labrador Sea Water of the eastern North Atlantic Ocean. *Deep-Sea Res. II* **46**, 165–179 (1999).
- Le Cann, B., Speer, K. A., Serpette, A., Paillet, J. & Reynaud, T. Lagrangian observations in the intergyre North-East Atlantic during the ARCANÉ and EUROFLOAT projects: Early results. *WOCE Int. Newsl.* **34**, 25–27 (1999).
- Curry, R. G. *A Database of Hydrographic Stations and Tools for Climatological Analysis* Technical Report WHOI-96-01 (Woods Hole Oceanographic Institution, Woods Hole, Massachusetts, 1996).
- McCartney, M. S. & Mauritzen, C. On the origin of the warm inflow to the Nordic Seas. *Prog. Oceanogr.* **51**, 125–214 (2001).
- McCartney, M. S. & Talley, L. D. Warm-to-cold conversion in the northern North Atlantic Ocean. *J. Phys. Oceanogr.* **14**, 922–935 (1984).
- Paillet, J., Arhan, M. & McCartney, M. S. Spreading of Labrador Sea water in the eastern North Atlantic. *J. Geophys. Res.* **103**, 10223–10239 (1998).
- Talley, L. D. & McCartney, M. S. Distribution and circulation of Labrador Sea water. *J. Phys. Oceanogr.* **12**, 1189–1205 (1982).
- Dickson, R. R. & Brown, J. The production of North Atlantic deep water: Sources, rates, and pathways. *J. Geophys. Res.* **99**, 12319–12341 (1994).
- Worthington, L. V. On the North Atlantic circulation. *Johns Hopkins Oceanogr. Stud.* **6**, 1–110 (1976).
- Pollard, R., Read, J. & Holliday, P. Circulation and mode waters of the North Atlantic subpolar gyre in 1996. *WOCE Int. Newsl.* **37**, 21–27 (1999).
- Krauss, W. The North Atlantic Current. *J. Geophys. Res.* **91**, 5061–5074 (1986).
- Rosby, T. The North Atlantic current and surrounding waters: At the crossroads. *Rev. Geophys.* **34**, 463–481 (1996).
- Belkin, I. M. & Levitus, S. Temporal variability of the Subarctic Front near the Charlie–Gibbs Fracture Zone. *J. Geophys. Res.* **101**, 28317–28324 (1996).
- Sy, A., Schauer, U. & Meincke, J. The North Atlantic Current and its associated hydrographic structure above and eastwards of the Mid-Atlantic Ridge. *Deep-Sea Res.* **39**, 825–853 (1992).
- Heywood, K. J., McDonagh, E. L. & White, M. A. Eddy kinetic energy of the North Atlantic subpolar gyre from satellite altimetry. *J. Geophys. Res.* **99**, 22525–22539 (1994).
- Iorga, M. C. & Lozier, M. S. Signatures of the Mediterranean outflow from a North Atlantic climatology, diagnostic velocity fields. *J. Geophys. Res.* **104**, 26011–26029 (1999).
- Pingree, R. D. & LeCann, B. Structure, strength and seasonality of the slope currents in the Bay of Biscay region. *J. Mar. Biol. Assoc. UK* **70**, 857–885 (1990).
- Bersch, M., Meincke, J. & Sy, A. Interannual thermocline changes in the northern North Atlantic. *Deep-Sea Res. II* **46**, 55–75 (1999).
- Bersch, M. NAO-induced changes of the upper-layer circulation in the northern North Atlantic Ocean. *J. Geophys. Res.* (in the press).
- Schmitz, W. J. Jr & McCartney, M. S. On the North Atlantic circulation. *Rev. Geophys.* **31**, 29–49 (1993).
- Rosby, H. T., Levine, E. & Connors, D. N. The isopycnal Swallow Float: A simple device for tracking water parcels in the ocean. *Prog. Oceanogr.* **14**, 511–525 (1985).
- Freeland, H. J. & Gould, W. J. Objective analysis of meso-scale ocean circulation features. *Deep-Sea Res.* **23**, 915–923 (1976).
- Zhang, H. M., Prater, M. D. & Rosby, T. Isopycnal Lagrangian statistics from the North Atlantic current RAFOS float observations. *J. Geophys. Res.* **106**, 13817–13836 (2001).

Acknowledgements We thank all the technicians and research assistants who contributed to this project for their help and support. We also thank H. Hunt Furey for help in the preparation of the figures. This work was supported by the US National Science Foundation, the Deutsche Forschungsgemeinschaft, the Institut Français de Recherche pour l'Exploitation de la Mer, the Centre National de la Recherche Scientifique, and the European Commission EUROFLOAT project.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.B. (e-mail: abower@whoi.edu).

A correlation between mid-ocean-ridge basalt chemistry and distance to continents

Eric Humler* & Jean Besse†

* Laboratoire des Geosciences Marines; and † Laboratoire de Géomagnétisme et de Palaeomagnétisme, IPGP-Université Denis Diderot, 4, Place Jussieu, 75252 Paris Cedex 05, France

To fully understand the structure and dynamics of the Earth's convecting mantle, the origins of temperature variations within the mantle need to be resolved. Different hypotheses have been proposed to account for these temperature variations: for example, heat coming from the decay of radioactive elements or heat flowing out of the Earth's core. In addition, theoretical studies^{1–5} suggest that the thermal properties of continental masses can affect mantle convection, but quantitative data that could allow us to test these models are scarce. To address this latter problem, we have examined the chemistry of mid-ocean-ridge basalt—which reflects the temperature of the source mantle^{6,7}—as a function of the distance of the ridge from the closest continental margin. No correlation is observed for oceanic ridges close to subduction zones or hotspots; subduction zones probably inhibit thermal transfer between the mantle beneath continents and ocean, whereas hotspots influence the major-element chemistry of ridge basalts, which makes their interpretation with respect to mantle temperature more difficult. However, we do observe a significant correlation for mid-oceanic basalts from the Atlantic and Indian oceans. From this, we conclude that the location of continental masses relative to active ridges influences the large-scale thermal structure of the mantle and we estimate that the mantle cools by 0.05 to 0.1 °C per kilometre from the continental margins.

Major-element data for oceanic basalts are important because they can be quantitatively linked to mantle temperatures, axial depth of oceanic ridges, crustal thickness and seismic tomography^{6–9}. Quantification of potential mantle temperatures based on basalt chemistry, experimental petrology and theories of melt extraction are now available^{6–7,10–12}. In particular, $\text{Na}_{8,0}$ (Na_2O content corrected to 8% MgO) can be used as a proxy for the extent of melting, whereas $\text{Fe}_{8,0}$ is a proxy for the mean pressure of melting. Using the Langmuir *et al.*⁷ formulation for mantle melting, basalts from shallow ridges (Kolbeinsey) with low $\text{Na}_{8,0}$ and high $\text{Fe}_{8,0}$ correspond to a solidus mantle temperature of about 1,460 °C. Deep ridges, such as the South West Indian Ridge (SWIR) with high $\text{Na}_{8,0}$ and low $\text{Fe}_{8,0}$ correspond to a solidus mantle temperature of about 1,210 °C.

The chemical composition of ancient crust can be used to determine palaeobathymetry of ancient ridge axes¹³. This approach has recently been applied to oceanic crust older than 80 Myr. The results were previously interpreted to infer that mantle temperatures during the Mesozoic were 50 °C hotter than today¹⁴. Such a large temperature decrease in such a short time interval cannot be reconciled with models of normal cooling of the Earth.

In the Earth, the thermal conditions at the upper boundary of the convecting mantle are not uniform. This is demonstrated in the most direct manner by the distribution of heat flow at the surface. Continental plates can act as insulating lids^{1–5,15,16} thereby imposing large-scale heterogeneities at the Earth's surface. These heterogeneities could play an important role in determining the Earth's overall thermal structure. Theoretical studies^{1–5} indicate that the amplitude of the process depends on the area, thickness and width of the continents, and on the duration of time during which they remain above a given area of the mantle before break-up. If continents do influence mantle temperatures, a correlation should exist between

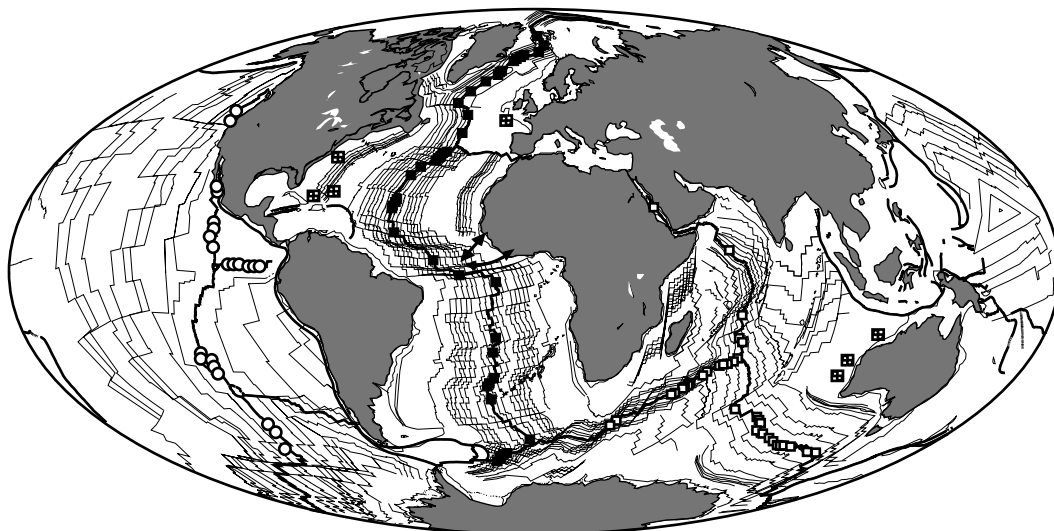


Figure 1 Location of samples used in this study. Open circles, Pacific Ocean; black squares, Atlantic Ocean; white squares, Indian Ocean and the Red Sea. Crossed squares are Ocean Drilling Program (ODP)-Deep Sea Drilling Project (DSDP) sites cored on oceanic crust older than 80 Myr in the Atlantic and Indian oceans (see ref. 14). The mean distances between ridge axes and continental margins (500 fathoms = 915 m) have been calculated for each site. Calculations use the shortest distance to continents (small arrow for all oceans) or distance along flow lines (long arrow only for the Atlantic). Up to

first order, oceanic ridges close to continents (Red Sea, Aden, Mid-Atlantic Ridge (MAR) at 69° N) are shallower than those far from continents (SWIR, close to Rodriguez Triple Junction, American–Antarctic discordance (AAD), MAR at the Equator). A complete data set with mid-ocean-ridge basalt chemistry and the associated measured distances to the nearby continental margins is available in the Supplementary Information (Tables A to D) or upon request to E.H.

the chemistry of ‘zero-age’ oceanic basalts (recording temperature variations within the mantle) and one or several of the parameters listed above. A rather obvious observation is that ridges close to continents are generally shallower than those far away from them. Because of the proposed link between mantle temperatures and ridge bathymetry^{6,7}, temperatures beneath ridges located close to a

continent should be hotter than mantle temperatures under oceanic ridges remote from continents.

Here, we propose a first attempt at relating the chemical composition of zero-age oceanic crust with distance to nearby continents. We have compiled published^{6,7,17} and unpublished¹⁸ major-element analyses for zero-age oceanic crust from the Atlantic,

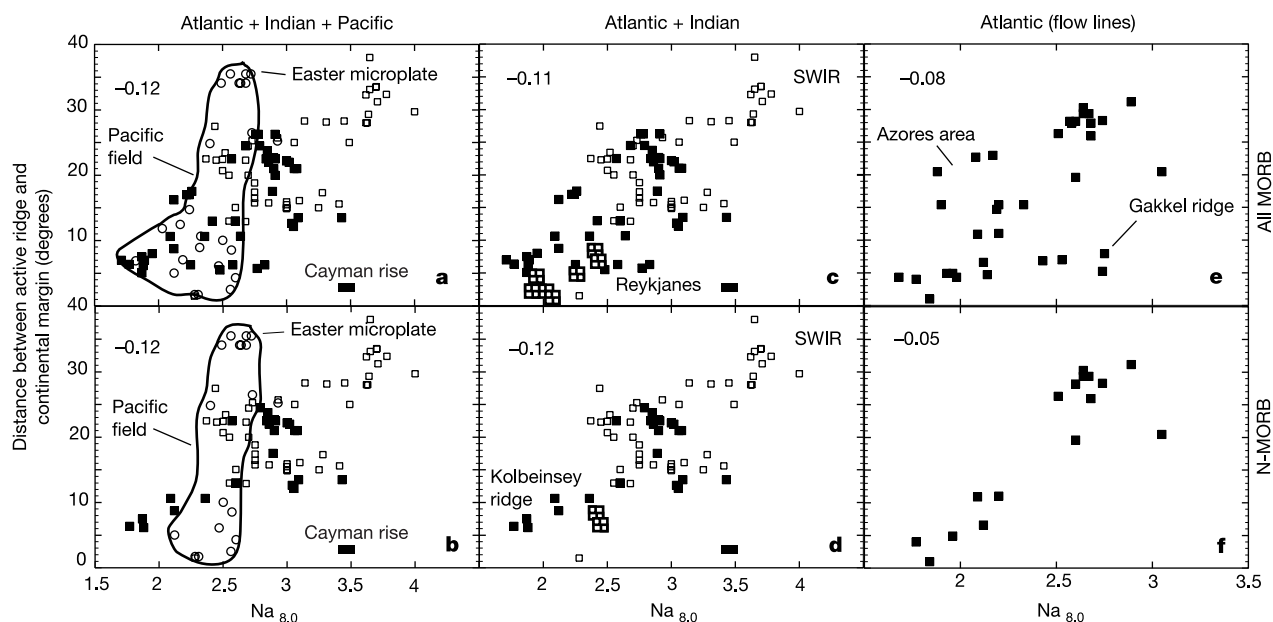


Figure 2 Covariation of distance (in degrees) of oceanic ridges from continental margins versus mid-ocean-ridge basalt (MORB) chemistry ($Na_{8.0}$) for the sites shown in Fig. 1. The Pacific field has been outlined; 1 degree = 111 km. **a, c, e**, All MORB have been used. **b, d, f**, Only MORB far from hotspots and hotspot margins have been used. Numbers in the upper left corner of the panel are the calculated ‘cooling gradients’ (see text) in $^{\circ}C\ km^{-1}$. **a**, A weak positive trend is obscured by data from the Pacific that do not show any correlation. **b**, Data screened for hotspot margin sites. **c**, Data are from Atlantic and Indian

oceans only. SWIR, South West Indian Ridge. **d**, Covariation between the mean distance (in degrees) of the oceanic ridges to continental margins versus $Na_{8.0}$ restricted to normal MORB (N-MORB) from the Atlantic and the Indian oceans. **e** and **f**, All MORB and N-MORB from the Atlantic ocean (ref. 6) between 75° N and 40° S. Basalts from Gakkel ridge fall off the trend (data shifted toward higher $Na_{8.0}$ than those expected from the general correlation). Samples close to the Azores are generally shifted toward low $Na_{8.0}$ values. Symbols as in Fig. 1.

Indian and Pacific oceans. The extended data set used in our study is representative of the global trends defined for ocean-ridge basalt chemistry in refs 6 and 7. For each site, we have calculated the distance between the ridge axis and the nearest continental margin defined by the 500-fathom isobath (see Tables A to C and Fig. A in the Supplementary Information).

Maps of the global ridge system show substantial distance variations of oceanic ridges from continental margins (Fig. 1). In the Atlantic, the distance between the active ridge axis and the continental margins varies from about 330 km at latitude 69° N to 3,000 km at latitude 37° S, that is, one order of magnitude. The minimum distance of Iceland from the north European continental margin is about 930 km, and from the North American continental margin, about 350 km. These distances are short compared to those measured between the SWIR axis at 64–68° E and the Antarctic or African continental margins (4,300 and 3,000 km respectively). In the Pacific, ridges are surrounded by subduction zones and thus the distance to a continent is equivalent to the distance to a subduction zone. These distances have been measured relative to both South and North America and, in the southern Pacific, to Antarctica.

The chemical composition of oceanic basalts is plotted against the mean distance to the continents in Fig. 2. The data show a broad positive trend in each diagram. Low $\text{Na}_{8,0}$ (that is, hot mantle) basalts (Kolbeinsey, Reykjanes ridges) are associated with small distances from the continental margins and high $\text{Na}_{8,0}$ (that is, cold mantle) basalts (SWIR) with large distances from the continents. Reverse correlations between $\text{Fe}_{8,0}$ and distance from the continents have been obtained (not shown in Fig. 2), suggesting a higher mean pressure of melting for oceanic ridges close to continents relative to those far from continents.

However, several points deviate substantially from the general trend. The corresponding sites are located in the Pacific (Easter microplate) and in the northern part of the East Pacific Rise (Fig. 2a and b). In the Atlantic (Fig. 2a, c and e), most of these sites are clustered between 71.5 and 74° N latitude (Gakkel ridge) and on the Cayman rise.

Some of the basalts used in our study are known to be ‘abnormal’. For instance, those from ridge segments which have been influenced by certain hotspots (the Azores, Jan Mayen, the Galapagos islands) have low $\text{Fe}_{8,0}$ and $\text{Na}_{8,0}$ (ref. 7) coupled with a low Sm/Yb_N for a given $\text{Ce}/\text{Sm}_N^{14}$ (here subscript N indicates normalized to chondrites). For those sites, major-element chemistry cannot be used to calculate a reliable mantle temperature. These sites have been removed in Fig. 2b, d and f. There is no chemical basis to screen out the data from the Pacific (except for those close to the Galapagos and Easter island; Fig. 2b). Some Pacific data do not fit the correlation (as mostly defined by the Atlantic and Indian Ocean data in Fig. 2d), probably because subduction inhibits thermal transfer between continent and ocean. This suggestion is supported by the fact that data fall on the trend (defined by the Atlantic and Indian oceans) in the south Pacific where subduction is weaker (present in the West but absent in the East). The possible effect of subduction zones on the uncorrelated data is also well illustrated in the case of the Cayman rise which is surrounded by the Puerto Rico and Middle America trenches. In the following, we restrict our analysis to normal mid-ocean-ridge basalt from the Indian and Atlantic basins, where subduction zones are less numerous than in the Pacific (Fig. 2d).

At first glance, based on Fig. 2c or d, it would seem that mantle temperatures beneath the ridge axis are regionally controlled by distances to continents. Here, we have termed the observed temperature decrease with distance to a continent a ‘cooling gradient’. The calculated cooling gradient (based on the correlations in Fig. 2) is about $0.1^\circ\text{C km}^{-1}$. In Figs 2e and f, distances between active ridges and continental margins have been measured (only for data from the Mid-Atlantic Ridge) along kinematic flow lines between

75° N and 40° S. The correlation improves considerably for basalts remote from hotspots and hotspot margins (Fig. 2f). The calculated cooling gradient in this case is about $0.05^\circ\text{C km}^{-1}$.

The chemical composition of older oceanic crust has previously been used to infer that mantle temperatures in the Mesozoic were on average 50°C hotter than today¹⁴. On the basis of plate reconstructions¹⁹, the distances to continents have been calculated in a hotspot reference frame and the data reported on Fig. 2c and d (crossed squares). These agree with the general correlation and fall close to the low $\text{Na}_{8,0}$ and short-distance endmember, suggesting that the apparent higher mean mantle temperature is linked to a past location of continents (see Table D in the Supplementary Information).

The theory of plate tectonics has been very successful in accounting for the velocity field of the plates at the Earth’s surface and in explaining large-scale geological phenomena such as subduction, continental collision, volcanism and seismicity. Studies of mantle convection have focused on observations in the oceans. In such studies, continents have often been treated as ‘aged’ oceanic lithosphere, passively following plate motions. The results obtained here, based on available chemical data, corroborate the hypothesis, based on studies of the dynamics of mantle convection, that continents affect the large-scale thermal structure of the mantle^{1–5,15,16}. □

Received 2 July; accepted 6 August 2002; doi:10.1038/nature01052.

- Olson, P. & Corcos, G. M. A boundary layer model for mantle convection with surface plates. *Geophys. J. R. Astron. Soc.* **62**, 195–219 (1980).
- Gurnis, M. Large-scale mantle convection and the aggregation and dispersal of supercontinents. *Nature* **332**, 695–699 (1988).
- Lowman, J. P. & Jarvis, G. T. Mantle convection flow reversals due to continental collisions. *Geophys. Res. Lett.* **20**, 25485–25497 (1993).
- Zhong, S. & Gurnis, M. Dynamic feedback between a continent-like raft and thermal convection. *J. Geophys. Res.* **98**, 12219–12232 (1993).
- Guillou, H. & Jaupart, C. On the effect of continents on mantle convection. *J. Geophys. Res.* **100**, 24217–24238 (1995).
- Klein, E. M. & Langmuir, C. H. Global correlations of ocean ridge basalt chemistry with axial depth and crustal thickness. *J. Geophys. Res.* **92**, 8089–8115 (1987).
- Langmuir, C. H., Klein, E. M. & Plank, T. Petrological systematics of mid-ocean ridge basalts: Constraints on melt generation beneath oceanic ridges. *AGU Geophys. Monogr.* **71**, 183–280 (1992).
- Humler, E., Thiriot, J. L. & Montagner, J. P. Global correlations of mid-ocean ridge basalt chemistry with tomographic images. *Nature* **364**, 225–228 (1993).
- Langmuir, C. H. Deep thoughts on the mantle. *Nature* **364**, 191–192 (1993).
- Albarède, F. How deep do common basaltic magmas form and differentiate? *J. Geophys. Res.* **97**, 10997–11009 (1992).
- McKenzie, D. & Bickle, M. J. The volume and composition of melt generated by extension of the lithosphere. *J. Petrol.* **25**, 625–679 (1988).
- Niu, Y. & Batiza, R. An empirical method for calculating melt compositions produced beneath mid-ocean ridges: application for axis and off-axis (seamounts) melting. *J. Geophys. Res.* **96**, 21753–21777 (1991).
- Keen, M. J., Klein, E. M. & Melson, W. G. Ocean ridge basalt compositions correlated with palaeobathymetry. *Nature* **345**, 423–426 (1990).
- Humler, E., Langmuir, C. H. & Daux, V. Depth versus age: new perspectives from the chemical compositions of ancient crust. *Earth Planet Sci. Lett.* **173**, 7–23 (1999).
- Elder, J. *The Bowels of the Earth* 222 (Oxford Univ. Press, Oxford, 1976).
- Anderson, D. L. Hot spots, polar wander, Mesozoic convection and the geoid. *Nature* **297**, 391–393 (1982).
- Schilling, J. G. *et al.* Thermal structure of the mantle beneath the equatorial Mid-Atlantic Ridge: inferences from the spatial variation of dredged basalt compositions. *J. Geophys. Res.* **100**, 10057–10079 (1995).
- Humler, E., Meyzen, C., Ludden, J. & Mevel, C. Geochemical variations in basaltic glasses from the South West Indian Ridge (49 to 69°E). *Eos* **79**, 878 (1998).
- Müller, D. M., Royer, J. Y. & Lawver, L. A. Revised plate motions relative to the hotspots from combined Atlantic and Indian Ocean hotspot tracks. *Geology* **21**, 275–278 (1993).

Supplementary Information accompanies the paper on Nature’s website (<http://www.nature.com/nature>).

Acknowledgements We thank C. Jaupart, A. Bonneville, L. E. Ricou, S. Labrosse, S. Gilder, P. Machetel and C. Sotin for discussions. We also thank V. Courtillot and M. Perfit for their comments and corrections, and E. Klein for comments on an earlier version of the manuscript. We used Paleomac software provided by J. P. Cogne (available at <http://www.ipgp.jussieu.fr/~cogne>).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to E.H. (e-mail: humler@ipgp.jussieu.fr).

General patterns of taxonomic and biomass partitioning in extant and fossil plant communities

Brian J. Enquist^{*†}, John P. Haskell[‡] & Bruce H. Tiffney[§]

^{*} Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85721, USA

[†] Center for Applied Biodiversity Research, Conservation International, 1919 M Street NW, Suite 600, Washington DC 20036, USA

[‡] Department of Fisheries and Wildlife and the Ecology Center, Utah State University, Logan, Utah 84322-5210, USA

[§] Department of Geology, University of California Santa Barbara, Santa Barbara, California 93106, USA

A central goal of evolutionary ecology is to identify the general features maintaining the diversity of species assemblages^{1–3}. Understanding the taxonomic and ecological characteristics of ecological communities provides a means to develop and test theories about the processes that regulate species coexistence and diversity. Here, using data from woody plant communities from different biogeographic regions, continents and geologic time periods, we show that the number of higher taxa is a general power-function of species richness that is significantly different from randomized assemblages. In general, we find that local communities are characterized by fewer higher taxa than would be expected by chance. The degree of taxonomic diversity is influenced by modes of dispersal and potential biotic interactions. Further, changes in local diversity are accompanied by regular changes in the partitioning of community biomass between taxa that are also described by a power function. Our results indicate that local and regional processes² have consistently regulated community diversity and biomass partitioning for millions of years.

Three broad categories of processes have been hypothesized to regulate local species richness across ecological communities: ecological interactions, phylogenetic history and historical contingency^{1–6}. Their relative influence on communities has been the subject of debate^{6–11}, but singly or in combination, these processes ultimately determine the composition of ecological communities and how community biomass is divided into discrete systematic

units. Here we use two large woody plant data sets to identify the relative importance and generality of the ecological, evolutionary and historical processes shaping the taxonomic structure and biomass division within local plant communities across the world.

We used 227 standardized one-tenth hectare samples of woody plant communities from around the world compiled by the late A.H. Gentry and the Missouri Botanical Garden^{9,10}. The observed relationship between species and higher taxa is not a constraint envelope, as would be expected if certain taxa sometimes dominated locally species-rich communities (Fig. 1). Instead, our analysis shows that the relationships between the number of species, S , the number of genera, G , and the number of families, F , are described by simple power-functions, where $F \propto S^{0.68}$ and $G \propto S^{0.941}$ (Fig. 2; see also refs 4, 12 and 13). These exponents are statistically invariant across and within major biogeographic provinces, continents and physiognomic types (see Table 1a in the Supplementary Information). Furthermore, taxonomic exponents for palaeocommunities are indistinguishable from those of extant communities (Fig. 3 and Supplementary Information). As species richness increases, congeneric species and confamilial genera become proportionately more common because the observed exponents between species and higher taxa are less than 1. As a result, an increase in local diversity is paralleled by a proportional increase in taxonomic similarity between taxa within the community.

The consistent macroecological nature of the relationship between the number of species and higher taxa across continents supports the assertions of some palaeontologists and conservationists that the number of higher taxa can be used as a predictor of species richness within local assemblages^{12–15} (see also ref. 16), although the relationship is not linear. Data from diverse taxa, including marine molluscs¹³, birds, mammals (J.P.H., unpublished work) and ants (M. Weiser, personal communication) suggest that local communities of animals are also characterized by approximate power-functions. The presence of general functions that describe the partitioning of subunits into higher taxonomic groups is of importance to ecologists and palaeobiologists requiring quick and robust estimates of local taxonomic richness in both present and historical communities^{12,13} and to theoretical and empirical ecologists studying the causes and consequences of biological diversity.

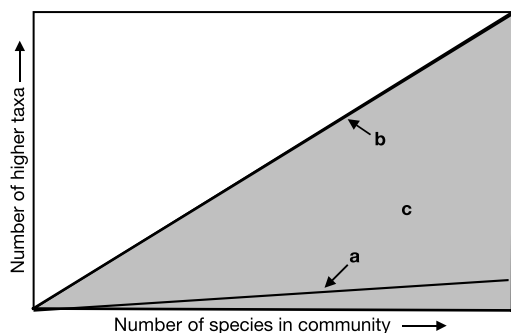


Figure 1 Three graphical hypotheses for the relationship between species richness and number of higher taxa within a local community. **a**, A positive relationship with a shallow slope: as species richness increases, the number of higher taxa (that is, genera or families) increases at a slower rate; additional species come from within an increasingly limited subset of higher taxa. **b**, A slope of unity represents the upper constraint boundary where increases in species richness result solely from the addition of higher taxa. **c**, The shaded area represents a constraint region in which the variance in abundance of higher taxa could be expected to increase with species richness: actual higher taxa abundance would be effectively unpredictable from species richness.

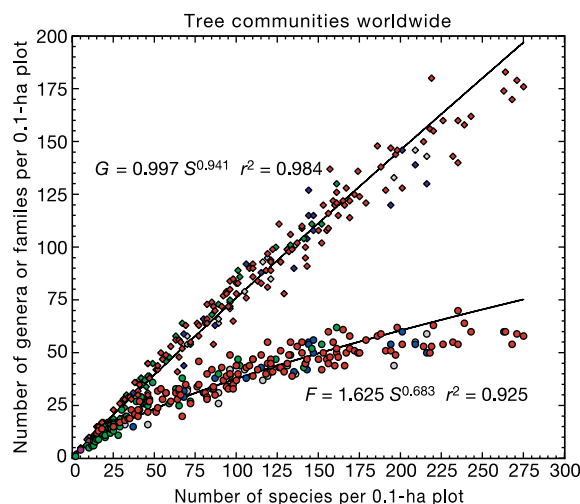


Figure 2 Relationship between the number of species and higher taxa across 227 0.1-ha sites from around the world. Species richness at individual sites ranges from 2 to 275. Diamonds, number of genera; circles, number of families. Colours correspond to samples from each continent (red, South America; blue, Asia; grey, Africa; green, North America; purple, Europe). Significant differences were not observed between the individual continents, hemispheres, and the world (see Table 1a,b in the Supplementary Information). G , number of genera; S , number of species; F , number of families.

To determine whether the local-scale taxonomic structure was significantly different from random samples of species pools of differing sizes, we constructed two Monte Carlo simulations. In all tests, assuming equal probability of colonization, the implicit null hypothesis is that the taxonomic structure of local communities is not distinguishable from a random sample of the species pool in which they were contained^{17,18} (see Methods and Supplementary Information). Randomized communities assembled from global, regional, continental and local scales are characterized by significantly higher exponents than observed (Table 1b,c in the Supplementary Information and Fig. 4). For a given species richness, 'real' communities consist of fewer numbers of genera and families than communities randomly assembled from species pools of different sizes^{4,5,18}. This accords with Elton's original observation⁴, that local communities are more taxonomically similar than expected by chance. With increasing species richness, generic richness rises at a proportionately slower rate and familial richness rises at an even slower rate (for example, $dF/dS \propto S^{-0.26}$ and $dG/dS \propto S^{-0.06}$). If, during evolutionary diversification, traits associated with physiology, morphology and ecology are conserved within a taxon, then congeners (members of the same genus) are expected to be more ecologically similar and compete more intensely with one another than noncongeners^{4-7,11,19}. As such, our observations are consistent with previous assertions¹⁹⁻²² that an increase in community species richness is accompanied by both (1) an increase in the proportional degree of ecological similarity between species due to taxonomic exponents less than 1 (see refs 2, 3, 5 and 23), and (2) an overall increase in total morphological or character diversity due to a uniform increase in higher taxa^{11,19,21-24}, although at an ever-slowing rate.

Two fundamental processes are likely to explain why generic and familial richness do not rise as quickly with increases in species numbers in natural communities as they do in our null models. First, taxa are likely to vary widely in their ability to gain access to local sites (that is, poor mixing of taxa within a given species pool owing to dispersal limitation^{2,19}). Second, biotic and abiotic features of a local site will influence the ability of taxa to colonize and survive within the site¹⁹. Patterns of residual variation (Fig. 1) support both of these hypotheses. For example, there is a significant negative residual correlation in generic and familial richness with local annual precipitation (genera: $r^2 = -0.155$, $n = 91$, $F = 16.33$, $P = 0.0001$; families: $r^2 = -0.148$, $n = 91$, $F = 15.51$, $P = 0.0002$), while there is a significant positive residual correlation for familial diversity and elevation (genera: $r^2 = 0.005$, $n = 145$, $F = 0.757$, $P = 0.386$; families: $r^2 = 0.101$, $n = 145$, $F = 15.98$, $P < 0.0001$). Thus, for a given species diversity, more dry and high-elevation sites have proportionally greater numbers of higher

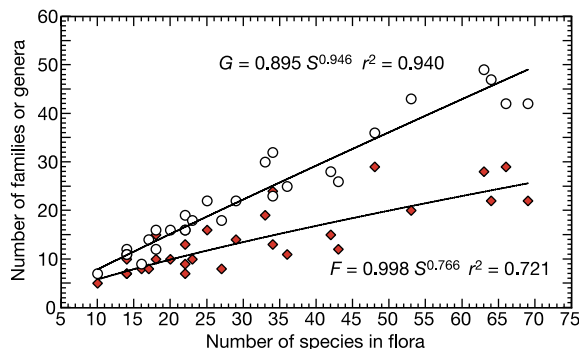


Figure 3 Relationship between the number of species and higher taxa across 28 local palaeoflora sites ranging from 4.5 to 45 Myr ago. Each point represents a single sampling locality (see ref. 10 and Supplementary Information). Open circles, the number of genera for a given number of species; filled diamonds, the number of families per species. Species richness at individual sites ranges from 10 to 69.

taxa than mesic low-elevation sites. This finding highlights the uniqueness of diverse high-elevation and dry forests^{9,10}. There is, however, no significant correlation between residual variation in generic or familial diversity with variation in local stem density, maximum stem diameter, total community biomass and absolute latitude. Our observations accord with a recognized association between wind/long-distance dispersal with (1) local rainfall and (2) the degree of insularity/increasing elevation^{25,26}. Thus, as indicated by our null model, deviation from the general taxonomic diversity function between sites (Fig. 1) seems to be significantly influenced by dispersal probabilities of a local flora.

Analysis of taxonomic exponents across size classes (a proxy for plant age) suggests a slight but significant decrease in the degree of taxonomic and hence ecological similarity as individuals within the community mature. By comparing the fitted regression parameters of the largest one-third of individuals with those of the smallest one-third we see a slight increase in the taxonomic exponent for both genera and families. The generic scaling exponent from the largest individuals ($b = 0.979 \pm 0.023$) falls outside the 95% confidence interval predicted by the smallest individuals ($b = 0.942 \pm 0.001$). The magnitude of the change for families is less than that observed for genera (largest $b = 0.776 \pm 0.032$; smallest $b = 0.754 \pm 0.020$). We cannot assess whether a greater proportion of smaller-sized taxa generates this directional change in the taxonomic exponent. However, the observed directional shift agrees with the prediction that disproportionate competitive thinning of related taxa influences local richness^{1,4,6}. The observed patterns of residual variation in taxonomic diversity are in accord with interpretations that the probabilities of local species establishment and extinction^{19,21-24,27} are governed in part by dispersal ability and the degree of ecological similarity of the interacting taxa, both of which are mediated by the local abiotic environment^{5-8,18-20}.

Our analyses exclude the possibility that the taxonomic structure influences the total biomass within natural assemblages¹⁹⁻²¹. As the total standing aboveground dry biomass, M_{tot} , is approximately constant with local richness ($M_{\text{tot}} \propto S^0$), increasing species richness within natural communities results in a finer division of biomass between taxa rather than an increase in total biomass^{10,28}. Instead, observed patterns of biomass partitioning are consistent with zero-sum models of community assembly (see refs 2 and 28) where a change in local richness is compensated for by a concomitant

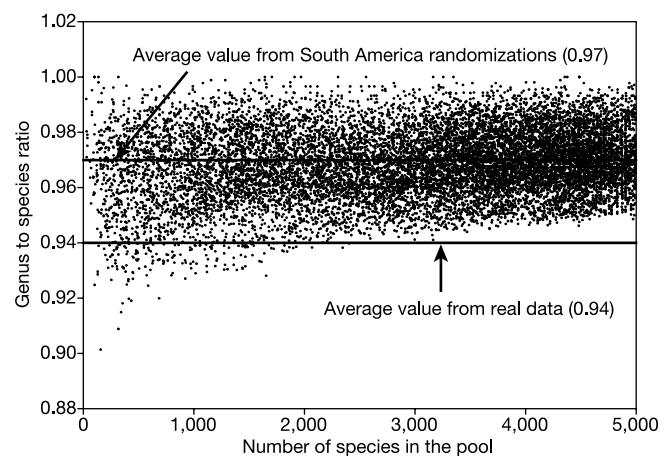


Figure 4 Relationship between species pool size and the results of randomization experiments using the total number of sampled sites from South America. The expected exponent describing the number of genera per species is relatively constant above a small pool size; and, except for the smallest pool sizes, it is consistently greater than the observed exponent. These results indicate that fewer genera (and families) comprise real world communities than would be expected by chance and that this result is largely independent of species pool size.

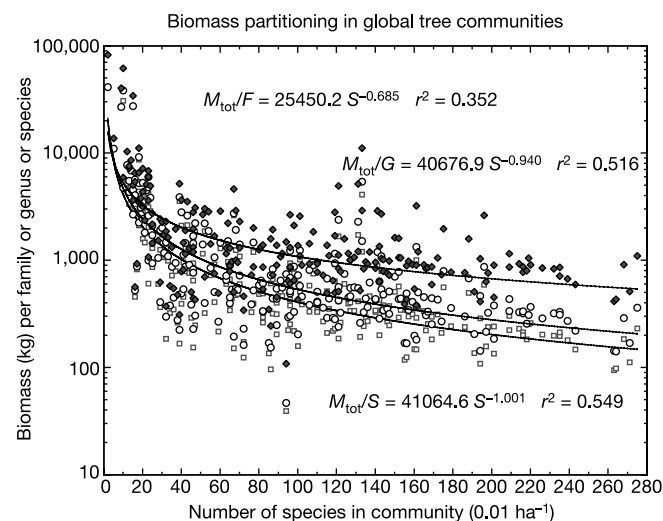


Figure 5 Biomass partitioning between species (open squares), genera (open circles), and families (filled diamonds) across global woody plant communities. The fitted inverse power-functions indicate a decrease in biomass per taxon with increasing richness. Although the relationships are statistically significant, patterns of residual variation indicate potential influences of other environmental and/or historical features on community biomass partitioning. M_{tot} , total standing aboveground dry biomass.

change in the amount of biomass contained within a given taxon. Thus, within a given community, the amount of biomass partitioned per family, M_{tot}/F , decreases with increasing species richness, where $M_{\text{tot}}/F \propto S^0/S^{0.683} \propto S^{-0.683}$. Similarly, the amount of local community biomass partitioned per genus and species must decrease as $M_{\text{tot}}/G = S^{-0.94}$ and $M_{\text{tot}}/S = S^{-1.0}$ respectively (Fig. 5).

Our results suggest that ecological communities have phylogenetic structure¹¹, reflected in the fact that general taxonomic and biomass partitioning functions describe species richness in woody plant communities. The presence of a power-function probably identifies the hierarchical and self-similar nature of phylogenetic diversification across a pool of species as reflected by systematic rankings² (see also ref. 16). We have used these empirical observations as a functional baseline by which to test null models of local composition and the importance of dispersal syndromes and biotic interactions.

The presence of a similar power-function that accounts for over 90% of the variation in taxonomic structure across contemporary and fossil woody plant communities suggests that the processes that promote and/or limit ecological similarity within communities (1) operate in a regular manner across broad geographical gradients, (2) reflect the ecological and evolutionary processes that dictate local composition, (3) uniquely quantify their net influence on community taxonomic structure, and (4) have probably operated in a consistent fashion across woody plant communities for millions of years. Our conclusions do not eliminate important roles for environment, history and other factors in shaping regional floras composed of many local communities, but they do suggest that these features are secondary in their influence on the patterns of local species coexistence. □

Methods

Data analysis

We analysed plant community data from a global data set collected by A.H. Gentry^{9,10}. The Gentry data set is remarkable in its coverage and uniform standardization. The data set includes 227 sites across six continents of tropical and temperate, closed canopy, forest communities ranging between 60.4° N to 40.43° S latitudes and elevations between 20 m and 3,050 m. Within each site, all plants with stem diameters of 2.5 cm at breast height (d.b.h.), including lianas, were sampled along ten 2 m × 50 m transects, totalling 0.1 ha at each site. Across sites, species diversity and number of individuals ranged between 2 and 275 and 52 and 1,005, respectively. The complete data set contains a total of 83,121

individual plants. Because the sample area is constant and the number of individuals per plot is approximately constant (see ref. 20), concerns related to sampling biases caused by differences in sample size or numbers of individuals are not important. Additional information is available through <http://www.mobot.org/MOBOT/Research/gentry/transect.shtml>. To test whether patterns observed in extant woody plant communities are consistent with those observed in geologic time, we compiled a similar database summarizing 29 floral samples ranging from 4.5 to 45 Myr ago from western North America (see Supplementary Information).

Statistical analysis

We first reassigned species randomly with equal sampling probabilities and with replacement to communities 1,000 times from a species pool composed of species from the world, eastern and western hemispheres, and individual continents. The number of species at each site was held constant. The log number of higher taxa was then regressed against the log number of species, just as in the original analysis. To preserve biological reality and to facilitate comparisons with the ratios from individual sites, the intercepts of the regressions were constrained to equal 0 ([1,1] in arithmetic space). In all but one case, the resulting average simulated exponents are larger than that observed in the real data (see Table 1 of the Supplementary Information). Only in the family-level analysis for Europe, where the sample size is extremely small ($n = 5$), was the average randomized exponent larger than the observed exponent at $P = 0.05$. Second, for South America we used a slightly more sensitive test by examining the ratios of the log number of higher taxa to the log number of species at individual sites. Specifics on this second test are listed in the Supplementary Information.

We also tested the effect of changing species pool size at individual sites in the South American data. This tests the effect of geographic scale on the taxonomic structure of the randomized communities. To quantify the effects of random reassignment and the size of the species pool, richness at each site was held constant, while the pool from which the species were drawn was increased sequentially by adding species from the next nearest sites, starting with only the nearest sites and working toward all sites on the continent. Again, species were drawn with equal probabilities within each spatial scale. Nearest-site distance, D , was calculated using a nearest euclidean neighbour distance algorithm: $D_{ij} = [(X_j - X_i)^2 + (Y_j - Y_i)^2]^{1/2}$ where X and Y represent the latitude and longitude of each site, i to j , respectively. Although the variance was influenced at extremely small pool sizes (that is, less than 500 species), that impact diminished quickly with pools composed of species from more than about 5–10 sites. We note that pool size is also an approximate measure of increasing geographic distance. The estimates of the exponents produced by the second analysis are plotted against the size of the species pool (Fig. 4). The result displays a curvilinear relationship between the randomized exponent and the size of the species pool but only at the smallest species pools.

Calculation of community and taxonomic biomass

We used empirical data reported in the literature from representative communities growing in latitudes and elevations equivalent to those communities in the Gentry data set to estimate above ground dry biomass from d.b.h. Biomass values are the same as those reported in ref. 28.

Received 18 December 2001; accepted 5 August 2002; doi:10.1038/nature01069.

- Hutchinson, G. E. Homage to Santa Rosalia, or why are there so many kinds of animals? *Am. Nat.* **93**, 145–159 (1959).
- Hubbell, S. P. *A Unified Theory of Biodiversity and Biogeography* (Princeton Univ. Press, Princeton, 2001).
- Ricklefs, R. E. & Schluter, D. (eds) *Species Diversity in Ecological Communities* (Univ. Chicago Press, Chicago, 1993).
- Elton, J. Competition and the structure of ecological communities. *Anim. Ecol.* **15**, 54–68 (1946).
- Williams, C. B. *Patterns in the Balance of Nature* (Academic, New York, 1964).
- Kinzig, A. P., Levin, S. A., Dushoff, J. & Pacala, S. Limiting similarity, species packing, and system stability for hierarchical competition-colonization models. *Am. Nat.* **153**, 371–383 (1999).
- MacArthur, R. H. & Levins, R. The limiting similarity, convergence, and divergence of coexisting species. *Am. Nat.* **101**, 377–385 (1967).
- Simberloff, D. S. Taxonomic diversity of island biotas. *Evolution* **24**, 23–47 (1970).
- Gentry, A. H. Changes in plant community diversity and floristic composition on environmental and geographic gradients. *Ann. Missouri Bot. Gard.* **75**, 1–34 (1988).
- Gentry, A. H. *Biological Relationships Between Africa and South America* (ed. Goldblatt, P.) 500–547 (Yale Univ. Press, New Haven, 1993).
- Webb, C. O. Exploring the phylogenetic structure of ecological communities: An example for rain forest trees. *Am. Nat.* **156**, 145–155 (2000).
- Williams, P. H., Humphries, C. J. & Gaston, K. J. Centers of seed-plant diversity—the family way. *Proc. R. Soc. Lond. B* **256**, 67–70 (1994).
- Roy, K., Jablonski, D. & Valentine, J. W. Higher taxa in biodiversity studies: patterns from eastern Pacific marine mollusks. *Phil. Trans. R. Soc. Lond. B* **351**, 1605–1613 (1996).
- Valentine, J. W. How many marine invertebrate fossil species? A new approximation. *J. Paleontol.* **44**, 410–415 (1970).
- Sepkoski, J. J. Jr *The Unity of Evolutionary Biology* (ed. Dudley, E. C.) 210–236 (Dioscorides, Portland, 1991).
- Robeck, H., Maley, C. C. & Donoghue, M. Taxonomy and temporal diversity patterns. *Paleobiology* **26**, 171–187 (2000).
- Manly, B. F. J. *Randomization, Bootstrap and Monte Carlo Methods in Biology* (Chapman & Hall, New York, 1997).
- Gotelli, N. J. & McCabe, D. J. Species co-occurrence: a meta-analysis of J.M. Diamond's assembly rules model. *Ecology* **83**, 2091–2096 (2002).
- MacArthur, R. H. & Wilson, E. O. *The Theory of Island Biogeography* (Princeton Univ. Press, Princeton, 1967).
- Darwin, C. *The Origin of Species* (John Murray, London, 1859).

21. Karr, J. R. & James, F. C. *Ecology and Evolution of Communities* (eds Cody, M. L. & Diamond, J. M.) 258–291 (Harvard Univ. Press, Cambridge, 1975).
22. Ricklefs, R. E. & O'Rourke, K. Aspect diversity in moths: a temperate-tropical comparison. *Evolution* **29**, 313–324 (1975).
23. Hubbell, S. P. *et al.* Light-gap disturbances, recruitment limitation, and tree diversity in a neotropical forest. *Science* **283**, 554–557 (1999).
24. Tilman, D., Lehman, C. L. & Thompson, K. T. Plant diversity and ecosystem productivity: Theoretical considerations. *Proc. Natl Acad. Sci. USA* **94**, 1857–1861 (1997).
25. Hughes, L. *et al.* Predicting dispersal spectra—a minimal set of hypotheses based on plant attributes. *J. Ecol.* **82**, 933–950 (1994).
26. Vazquez, J. A. G. & Givnish, T. J. Altitudinal gradients in tropical forest composition, structure, and diversity in the Sierra de Manantlan. *J. Ecol.* **86**, 999–1020 (1998).
27. Ribbens, E., Silander, J. A. & Pacala, S. W. Seedling recruitment in forests: Calibrating models to predict patterns of tree seedling dispersion. *Ecology* **75**, 1794–1806 (1994).
28. Enquist, B. J. & Niklas, K. J. Invariant scaling relations across tree-dominated communities. *Nature* **401**, 655–660 (2001).

Supplementary Information accompanies the paper on *Nature's* website
(<http://www.nature.com/nature>).

Acknowledgements We thank J. H. Brown, S. Collins, R. Colwell, M. J. Donoghue, N. J. Gotelli, D. Post, S. P. Hubbell, C. J. Humphries, W. P. Maddison, K. J. Niklas, N. Pittman, F. A. Smith, M. Weiser, J. Williams, R. Whittaker and the members of the NCEAS Body Size Working Group and Phylogenies and Community Ecology Working Group for critical discussions and/or comments on earlier drafts. In particular, C. O. Webb provided valuable comments. This work stems in part from the Body Size in Ecology and Evolution Working Group (F.A. Smith, Principal Investigator) sponsored by The National Center for Ecological Analysis and Synthesis (NCEAS, a national centre funded by the NSF, the University of California Santa Barbara and the State of California). B.J.E. was supported by the NSF and NCEAS. J.P.H. was supported by a student internship from NCEAS. Computer resources for the simulations were provided by the UNM Department of Biology, Sevilleta Long Term Ecological Research site and NCEAS.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to B.J.E. (e-mail: benquist@u.arizona.edu).

Genetic similarity between mates and extra-pair parentage in three species of shorebirds

Donald Blomqvist*, Malte Andersson†, Clemens Küpper*, Innes C. Cuthill‡, János Kis§, Richard B. Lanctot||, Brett K. Sandercock¶, Tamás Székely#, Johan Wallander† & Bart Kempnaers☆

* Konrad Lorenz Institute for Comparative Ethology, Austrian Academy of Sciences, Savoyenstrasse 1a, A-1160 Vienna, Austria

† Department of Zoology, Animal Ecology, Göteborg University, Box 463, SE-405 30 Gothenburg, Sweden

‡ Centre for Behavioural Biology, School of Biological Sciences, University of Bristol, Woodland Road, Bristol, BS8 1UG, UK

§ Behavioural Biology Research Group, Institute for Zoology, Faculty of Veterinary Science, Szent István University, POB 2, H-1400 Budapest, Hungary

|| US Fish and Wildlife Service, Migratory Bird Management, 1011 East Tudor Road, MS 201 Anchorage, Alaska 99503, USA

¶ Division of Biology, Kansas State University, Manhattan, Kansas 66505-4901, USA

Department of Biology and Biochemistry, University of Bath, Bath BA2 7AY, UK

☆ Max Planck Research Centre for Ornithology, PO Box 1564, D-82305 Starnberg (Seewiesen), Germany

Matings between close relatives often reduce the fitness of offspring, probably because homozygosity leads to the expression of recessive deleterious alleles^{1–5}. Studies of several animals have shown that reproductive success is lower when genetic similarity between parents is high^{4–7}, and that survival and other measures of fitness increase with individual levels of genetic diversity^{8–11}.

These studies indicate that natural selection may favour the avoidance of matings with genetically similar individuals. But constraints on social mate choice, such as a lack of alternatives, can lead to pairing with genetically similar mates. In such cases, it has been suggested that females may seek extra-pair copulations with less related males⁴, but the evidence is weak or lacking^{4,5}. Here we report a strong positive relationship between the genetic similarity of social pair members and the occurrence of extra-pair paternity and maternity ('quasi-parasitism') in three species of shorebirds. We propose that extra-pair parentage may represent adaptive behavioural strategies to avoid the negative effects of pairing with a genetically similar mate.

Molecular studies of socially monogamous birds have shown that broods often contain offspring that are not related to one of the parents tending the nest¹². Extra-pair fertilizations can result from females engaging in copulations with extra-pair males (extra-pair paternity; EPP), or from males copulating with extra-pair females that lay their eggs in the male's nest (quasi-parasitism; QP). Generally, EPP is common in passerines (songbirds) and, although other hypotheses cannot be discarded, it may be explained by females seeking 'good genes' for their offspring^{12,13}. In contrast, EPP is less common in non-passerine birds¹⁴, in which its adaptive significance remains unexplained. QP is rare among birds and poorly understood¹⁵. Here we propose an adaptive explanation for the occurrence of EPP and QP in non-passerine birds and show that it is over-represented in pairs with genetically similar mates.

We examined genetic parentage in western sandpipers *Calidris mauri*, common sandpipers *Actitis hypoleuca* and Kentish plovers *Charadrius alexandrinus*. Multilocus DNA fingerprinting identified low rates of EPP in these birds (Table 1), comparable to those found in most other shorebirds (ref. 16, and references therein). In contrast to previous studies, however, we also found evidence for QP in two of the species examined (Table 1). Thus, our study species are predominantly genetically monogamous, with alternative reproductive behaviours occurring at low frequencies.

The females laying the quasi-parasitic eggs may have been either mated (and thus were having extra-pair copulations) or unmated (floaters). Because we did not identify extra-pair parents in our study, we do not have evidence that can directly separate these two possibilities. If quasi-parasites were mated, however, we would expect their parasitic eggs to be fathered by their social mate (unless those females have extreme control over paternity). Given the low rate of EPP and the lack of intraspecific brood parasitism (only two cases documented in the Kentish plover; C.K., unpublished data), it seems more likely that the quasi-parasites were floaters. Observations suggest that female floaters are present in the common sandpiper (M.A., unpublished data), but information for the other species is lacking. In all cases of QP the clutch sizes were not increased, suggesting that the parasitic female (or the receiving

Table 1 Frequency of extra-pair fertilizations in three species of shorebirds determined by DNA fingerprinting

Species	Number of broods (number of chicks)	EPFs % (n)	EPP % (n)	QP % (n)
Kentish plover	65 (170)	4.6 (3) 2.9 (5)*	1.5 (1) 0.6 (1)	3.1 (2) 1.2 (2)
Western sandpiper	25 (61)	8.0 (2) 6.6 (4)	8.0 (2) 6.6 (4)	0 (0) 0 (0)
Common sandpiper	15 (53)	20.0 (3) 7.5 (4)	6.7 (1) 1.8 (1)	13.3 (2) 5.7 (3)

Only broods where both putative parents were fingerprinted are included. EPF, extra-pair fertilization; EPP, extra-pair paternity; QP, quasi-parasitism.

*Two EPF chicks could not be classified as EPP or QP owing to high band-sharing between pair members.

male) removed one or two eggs from the host female's nest. Alternatively, the parasitic female may have laid her eggs early in the laying sequence of the host female and the latter stopped laying when the nest contained the normal number of eggs.

In all three species, the occurrence of extra-pair fertilizations (EPP and QP combined) was related to genetic similarity between social pair members, which was estimated as band-sharing from multilocus DNA fingerprints (Fig. 1; logistic regression; Kentish plover: $\chi^2(1) = 7.99$, $P = 0.0047$; western sandpiper: $\chi^2(1) = 13.9$, $P = 0.00019$; common sandpiper: $\chi^2(1) = 5.98$, $P = 0.014$). Because regression analysis is sensitive to outliers, we also tested the relationship using ranked data. This confirmed that band-sharing between the mates was significantly higher in pairs tending broods with extra-pair offspring than in pairs with only within-pair young (Fig. 1; Mann–Whitney U -test; Kentish plover: $U = 29.5$, $P = 0.046$; western sandpiper: $U = 0$, $P = 0.021$; common sandpiper: $U = 3$, $P = 0.027$). The combined probability, based on the results of the Mann–Whitney U -tests, was $P < 0.01$ ($\chi^2(6) = 21.1$). An independent scoring of band-sharing between mates, carried out

by a colleague who did not know whether broods contained extra-pair young or not (Methods), yielded a qualitatively similar result (combined Mann–Whitney tests; $\chi^2(6) = 15.3$, $P < 0.02$). Thus, genetic similarity between mates significantly predicted the occurrence of extra-pair fertilizations.

Our data are insufficient to examine EPP and QP separately for each species. But because the genetic analyses were carried out in the same laboratory and by the same protocols (Methods), we can pool the data after standardizing the band-sharing values for each species (by subtracting the species-specific mean and dividing by the species-specific standard deviation). The pooled data show that the genetic similarity between mates explains the occurrence of EPP and QP, independently (logistic regression, $n = 105$ broods in total; EPP (4 broods): $\chi^2(1) = 5.47$, $P = 0.019$; QP (4 broods): $\chi^2(1) = 11.2$, $P = 0.0016$). We obtained similar results using Mann–Whitney U -tests on ranked data (data not shown). Note that we conservatively included the broods with QP in the group 'no EPP' in the first logistic regression, and the broods with EPP in the group 'no QP' in the second test (excluding these broods gives $\chi^2(1) = 17.0$, $P = 0.00004$ for EPP and $\chi^2(1) = 11.2$, $P = 0.00081$ for QP).

We have shown that in three species of shorebirds, breeding in North America, Europe and Asia, respectively, parents tending broods with extra-pair young are genetically more similar than those rearing exclusively within-pair chicks. This suggests that, in these species, males and females are more likely to engage in extra-pair matings when they are closely related to their social mate. This behaviour should be adaptive if high genetic similarity between parents has negative fitness consequences for their genetic offspring, and if genetic similarity with the extra-pair mate is lower. We do not have sufficient data to test this in our species, but the first assumption has been verified repeatedly^{1–11} and several studies have suggested that EPP may function to avoid the negative effects of inbreeding^{4,17–19}.

We know of only two other studies that have directly examined the relationship between genetic similarity of social mates and the occurrence of extra-pair fertilizations. In great reed warblers *Acrocephalus arundinaceus*, four of five females shared fewer bands with their extra-pair mate than with their social mate⁴. By contrast, in blue tits *Parus caeruleus*, band-sharing between the social parents did not differ for broods with or without extra-pair young, and the genetic similarity between the female and her social mate did not differ from that between the female and her extra-pair mate ($n = 18$ comparisons)⁵. In both species, there is evidence that females seek extra-pair copulations with high-quality males to increase the genetic quality of their offspring (the 'good genes' hypothesis)^{20,21}.

With respect to QP, our data suggest that a male may copulate with a female other than his social mate and allow her to lay one or more eggs in his nest. This is in agreement with the hypothesis that QP is male driven¹⁵ and may be adaptive, increasing the genetic quality of his offspring (but see ref. 15 for other explanations for the occurrence of QP). Note that, from a male perspective, QP differs from extra-pair copulations in that, first, males involved in QP do not gain extra offspring and, second, they do provide care for the QP offspring.

If males or females adjust their choice of genetic mate as suggested by our results, then an intriguing implication is that they can assess the genetic similarity with their social mate. The band-sharing between parents that raised extra-pair offspring varied between 5.7% (common sandpiper) and 57.1% (Kentish plover), suggesting that their ability to recognize kin is not restricted to first-order relatives. In support of this possibility, a study of lekking peacocks *Pavo cristatus* suggests that males can assess their genetic similarity with other males independently of social learning and environmental cues²², and such ability may be more widespread^{23,24}. Alternatively, a bias against genetically similar mates may

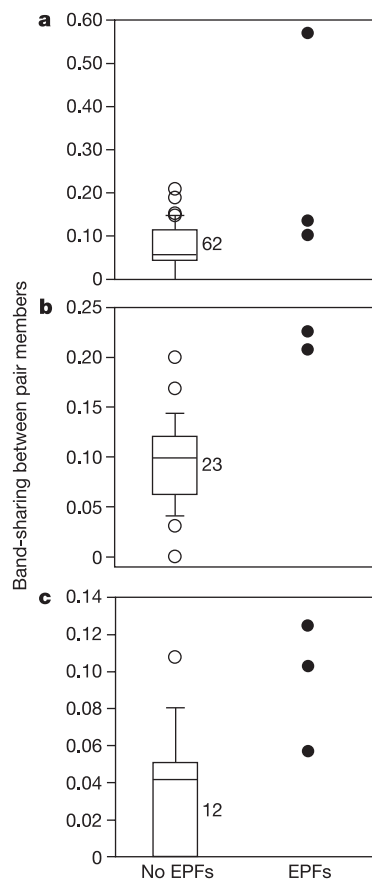


Figure 1 Band-sharing between mates and occurrence of extra-pair fertilizations (EPFs) in three species of shorebirds: **a**, Kentish plover; **b**, western sandpiper; and **c**, common sandpiper. EPFs include EPP and QP. For broods without extra-pair young, data are presented as box plots showing the 25th and 75th percentiles (box), median (line within the box), 10th and 90th percentiles (bars), and data points outside the latter percentiles. The number of broods is given beside the box. For broods with EPFs, the individual data points are shown. In all species, the difference in band-sharing between broods with or without extra-pair young was statistically significant (combined probability: $P < 0.01$, see text). The mean $\pm$ s.e.m. number of scored bands per adult was 18.8 ± 0.4 (Kentish plover, $n = 130$), 35.8 ± 1.0 (western sandpiper, $n = 50$) and 20.8 ± 1.0 (common sandpiper, $n = 30$).

arise through sperm competition if fertilization success is reduced for sperm that are genetically similar to the egg¹³, but this mechanism cannot explain the occurrence of QP. The cues used by birds to assess genetic similarity are unknown, presenting a challenging problem for future research.

In conclusion, we propose that EPP and QP in shorebirds and other non-passerine birds may be adaptive responses to avoid inbreeding depression or other negative effects of genetic similarity between social mates. □

Methods

Species and study sites

We studied western sandpipers at Nome, Alaska (64° 20' N, 164° 56' W), Kentish plovers at Tuzla, Turkey (36° 43' N, 35° 03' E) and common sandpipers at Sävå, Sweden (57° 47' N, 12° 19' E). All three species are waders or shorebirds (suborder Charadrii), have small clutches (3–4 eggs) and precocial young²⁵. Both sexes incubate the eggs, but males usually provide most care for the young. Western and common sandpipers are socially monogamous, whereas sequential polyandry is frequent in Kentish plovers²⁵. Breeding habitats consisted of tundra ponds and low ridges along the coast (western sandpiper), inland salt marshes (Kentish plover) and forested riverbanks (common sandpiper). We collected blood and tissue samples for fingerprinting from these populations in 1996 (western sandpiper), 1998–1999 (Kentish plover) and 1998–2000 (common sandpiper), respectively. Parents were caught while incubating or tending newly hatched chicks, whereas chicks usually were caught in or near the nest soon after hatching. For most pairs, observations from the pre-laying period are lacking. Thus, rapid mate replacement cannot be completely excluded as an alternative explanation for the occurrence of EPP.

Genetic analyses

We determined genetic parentage by multilocus DNA fingerprinting²⁶. Nuclear DNA was extracted from blood or tissue samples (dead chicks) using proteinase K and phenol/chloroform/isoamylalcohol. We separated 3–7 µg of *Hae*III-digested DNA on 0.8% agarose gels (20 × 40 cm) by electrophoresis at 1.2 V cm⁻¹ for 40 h. The DNA was transferred to nylon membranes using Southern blotting and hybridized with the multilocus probe *per*²⁷. The probe was radioactively labelled with [³²P]dCTP by random priming using the Prime-a-Gene labelling system (Promega).

Fingerprints were scored by standard methods²⁸ by C.K. (Kentish plover) and D.B. (common and western sandpiper). Scoring was done blind with respect to the tested hypothesis. We also obtained an independent analysis by asking a colleague (J. T. Lifjeld, Zoological Museum, University of Oslo, Norway) to score band-sharing²⁶ between mates, which he did without knowing whether their broods contained extra-pair young or not. We excluded a tending parent as a genetic parent when chick fingerprints showed several unattributable DNA fragments (novel bands) and low band-sharing with the parent in question. Pooling the three species in our study, extra-pair chicks showed 4–17 novel bands (6.9 ± 1.0 (mean ± s.e.m), *n* = 11) and shared 14.0–27.8% of the bands with the excluded parent (23.3 ± 1.2, *n* = 11). Band-sharing between non-excluded parents and their offspring varied between 32.0 and 78.8% (mother-offspring, 51.5 ± 0.6, *n* = 271) and between 31.3 and 74.5% (father-offspring, 52.0 ± 0.6, *n* = 271), respectively, with 0–2 novel bands owing to mutation or other random causes.

Measures of genetic similarity based on DNA fingerprinting are more reliable if it can be shown that bands segregate independently from parents to young following mendelian inheritance²⁶. We checked this by carrying out a segregation analysis on 15 randomly selected families (five in each species) without extra-pair young. In each family, we scored unique parental bands (males: 21.6 ± 2.0 bands, range 11–38; females: 21.7 ± 1.5, range 12–33) as either present (1) or absent (0) in their chicks. Mean transmission frequencies of single parental fragments were close to the expected value (0.50) for unlinked loci; they varied between 0.45 and 0.57 for paternal pedigrees (pooled mean = 0.51) and between 0.45 and 0.56 for maternal pedigrees (pooled mean = 0.50). In addition, only 10% of the parental bands (*n* = 649) were transmitted to all offspring in a brood, suggesting that most fragments were from heterozygous loci.

We examined linkage and allelism by calculating correlation coefficients (*r*) for all pairwise combinations of parental bands. Mean *r* varied between -0.03 and +0.14 in the paternal pedigrees (pooled mean 0.04) and between -0.04 and +0.14 in the maternal pedigrees (pooled mean 0.02), suggesting that most bands segregated independently. Indications of consistent co-segregation (*r* = 1.0) or allelism (*r* = -1.0) occurred infrequently in all families; on average, 6–9% (paternal pedigrees) and 7–8% (maternal pedigrees) of the pairwise combinations showed either of these patterns. Given the small brood sizes in our study species, however, several (if not all) of these cases may be coincidental. Consistently, the proportion of band combinations showing either co-segregation or allelism was significantly higher in families with three chicks than in those with four chicks (Mann-Whitney *U*-test combining both sexes: co-segregation, *U* = 63, *P* = 0.04, *n* = 16; allelism, *U* = 52, *P* = 0.01, *n* = 14). We therefore conclude that most bands stem from alleles at unlinked heterozygous loci, as found in many other studies using DNA fingerprinting (for example, see refs 5, 15).

In the field, adults were sexed by behaviour (such as courtship displays), size (especially bill length) or plumage characteristics (Kentish plovers)²⁵. To confirm cases of EPP and QP, we determined the sex of all parents with extra-pair offspring by polymerase chain reaction

(PCR) amplification of the *CHD-W* and *CHD-Z* genes using two sets of primers for each individual: P2 and P8 (ref. 29), and 3007 and 3112 (ref. 30), respectively. In brief, amplification was done in 10-µl reactions using a touchdown profile (Perkin Elmer 9600): 94 °C for 5 min; 60 °C, 72 °C, 94 °C, 58 °C, 72 °C, 94 °C, 52 °C and 72 °C for 30 s; 25 cycles at 94 °C, 50 °C and 72 °C for 30 s; and finally 4 °C for 10 min. We analysed the PCR products with an automatic sequencer (ABI 310 Genetic Analyzers, PE Biosystems), which in all cases confirmed the previously assigned sex.

Received 3 July; accepted 22 August 2002; doi:10.1038/nature01104.

- van Noordwijk, A. J. & Scharloo, W. Inbreeding in an island population of the great tit. *Evolution* **35**, 674–688 (1981).
- Thornhill, N. W. (ed.) *The Natural History of Inbreeding and Outbreeding* (Univ. Chicago Press, Chicago, 1993).
- Jiménez, J. A., Hughes, K. A., Alaks, G., Graham, L. & Lacy, R. C. An experimental study of inbreeding depression in a natural habitat. *Science* **266**, 271–273 (1994).
- Bensch, S., Hasselquist, D. & von Schantz, T. Genetic similarity between parents predicts hatching failure: non-incestuous inbreeding in the great reed warbler? *Evolution* **48**, 317–326 (1994).
- Kempnaers, B., Adriaenssens, F., van Noordwijk, A. J. & Dhondt, A. A. Genetic similarity, inbreeding and hatching failure in blue tits: are unhatched eggs infertile? *Proc. R. Soc. Lond. B* **263**, 179–185 (1996).
- Olsson, M. & Madsen, T. Promiscuity in sand lizards (*Lacerta agilis*) and adder snakes (*Vipera berus*): Causes and consequences. *J. Hered.* **92**, 190–197 (2001).
- Tregenza, T. & Wedell, N. Polyandrous females avoid costs of inbreeding. *Nature* **415**, 71–73 (2002).
- Amos, W. *et al.* The influence of parental relatedness on reproductive success. *Proc. R. Soc. Lond. B* **268**, 2021–2028 (2001).
- Coltman, D. W., Bowen, W. D. & Wright, J. M. Birth weight and neonatal survival of harbour seal pups are positively correlated with genetic variation measured by microsatellites. *Proc. R. Soc. Lond. B* **265**, 803–809 (1998).
- Coulson, T. N. *et al.* Microsatellites reveal heterosis in red deer. *Proc. R. Soc. Lond. B* **265**, 489–495 (1998).
- Hansson, B., Bensch, S., Hasselquist, D. & Åkesson, M. Microsatellite diversity predicts recruitment of sibling great reed warblers. *Proc. R. Soc. Lond. B* **268**, 1287–1291 (2001).
- Petrie, M. & Kempnaers, B. Extra-pair paternity in birds: explaining variation between species and populations. *Trends Ecol. Evol.* **13**, 52–58 (1998).
- Jennions, M. D. & Petrie, M. Why do females mate multiply? A review of the genetic benefits. *Biol. Rev.* **75**, 21–64 (2000).
- Westneat, D. F. & Sherman, P. W. Density and extra-pair fertilizations in birds: a comparative analysis. *Behav. Ecol. Sociobiol.* **41**, 205–215 (1997).
- Alves, M. A. S. & Bryant, D. M. Brood parasitism in the sand martin, *Riparia riparia*: evidence for two parasitic strategies in a colonial passerine. *Anim. Behav.* **56**, 1323–1331 (1998).
- Wallander, J., Blomqvist, D. & Lifjeld, J. T. Genetic and social monogamy—does it occur without mate guarding in the ringed plover? *Ethology* **107**, 561–572 (2001).
- Brooker, M. G., Rowley, I., Adams, M. & Baverstock, P. R. Promiscuity: an inbreeding avoidance mechanism in a socially monogamous species? *Behav. Ecol. Sociobiol.* **26**, 191–199 (1990).
- Stockley, P., Searle, J. B., Macdonald, D. W. & Jones, C. S. Female multiple mating behaviour in the common shrew as a strategy to reduce inbreeding. *Proc. R. Soc. Lond. B* **254**, 173–179 (1993).
- Brown, J. L. A theory of mate choice based on heterozygosity. *Behav. Ecol.* **8**, 60–65 (1997).
- Hasselquist, D., Bensch, S. & von Schantz, T. Correlation between male song repertoire, extra-pair paternity and offspring survival in the great reed warbler. *Proc. R. Soc. Lond. B* **266**, 813–816 (1999).
- Kempnaers, B. *et al.* Extra-pair paternity results from female preference for high-quality males in the blue tit. *Nature* **357**, 494–496 (1992).
- Petrie, M., Krupa, A. & Burke, T. Peacocks lek with relatives even in the absence of social and environmental cues. *Nature* **401**, 155–157 (1999).
- Höglund, J., Alatalo, R. V., Lundberg, A., Rintamäki, P. T. & Lindell, J. Microsatellite markers reveal the potential for kin selection on black grouse leks. *Proc. R. Soc. Lond. B* **266**, 813–816 (1999).
- Shorey, L., Piernney, S., Stone, J. & Höglund, J. Fine-scale genetic structuring on *Manacus manacus* leks. *Nature* **408**, 352–353 (2000).
- del Hoyo, J., Elliott, A. & Sargatal, J. (eds) *Handbook of the Birds of the World, Vol. 3. Hoatzin to Auks* (Lynx Edicions, Barcelona, 1996).
- Carter, R. E. *Molecular Methods in Ecology* (ed. Baker, A. J.) 113–135 (Blackwell Science, Oxford, 2000).
- Shin, H.-S., Bargiello, T. A., Clark, B. T., Jackson, F. R. & Young, M. W. An unusual coding sequence from a *Drosophila* clock gene is conserved in vertebrates. *Nature* **317**, 445–448 (1985).
- Westneat, D. F. Polygyny and extrapair fertilizations in eastern red-winged blackbirds (*Agelaius phoeniceus*). *Behav. Ecol.* **4**, 49–60 (1993).
- Griffiths, R., Double, M. C., Orr, K. & Dawson, R. J. G. A DNA test to sex most birds. *Mol. Ecol.* **7**, 1071–1075 (1998).
- Ellegren, H. & Fridolfsson, A.-K. Male-driven evolution of DNA sequences in birds. *Nature Genet.* **17**, 182–184 (1997).

Acknowledgements We thank J. Lifjeld for scoring fingerprints; D. Hasselquist, A. Johnsen, K. Lessells, B. Sheldon and R. Wagner for comments on the manuscript; and U. Bläsi, I. Moll and K. Carter for logistic support. The study was supported by the Konrad Lorenz Institute for Comparative Ethology, Vienna, and by grants from the Natural Environment Research Council UK, the Hungarian Scientific Research Fund and the Swedish Research Council.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to B.K. (e-mail: B.Kempnaers@erl.ornithol.mpg.de).

Attentional modulation in visual cortex depends on task timing

Geoffrey M. Ghose & John H. R. Maunsell

Division of Neuroscience and Howard Hughes Medical Institute, Baylor College of Medicine, Houston, Texas 77030, USA

Paying attention to a stimulus selectively increases the ability to process it. For example, when subjects attend to a specific region of a visual scene, their sensitivity to changes at that location increases. A large number of studies describe the behavioural consequences and neurophysiological correlates of attending to spatial locations^{1–8}. There has, in contrast, been little study of the allocation of attention over time^{9,10}. Because subjects can anticipate predictable events with great temporal precision^{11–15}, it seems probable that they might dynamically shift their attention when performing a familiar perceptual task whose constraints changed over time. We trained monkeys to respond to a stimulus change where the probability of occurrence changed over time. Recording from area V4 of the visual cortex in these animals, we found that the modulation of neuronal responses changed according to the probability of the change occurring at that instant. Thus, we show that the attentional modulation of sensory neurons reflects a subject's anticipation of the timing of behaviourally relevant events.

Two rhesus monkeys were trained to maintain fixation on a small dot and release a lever when a stimulus at a cued peripheral location changed orientation. The monkeys were required to ignore orientation changes of stimuli presented simultaneously at one to three other locations (Fig. 1a). The physiological effects of spatial attention were assessed by comparing responses from individual neurons between trials in which the visual stimulation was identical, but the cued location was different. Orientation changes occurred at random times after stimulus onset, but the probability of a change occurring as a function of time within each trial was fixed (Fig. 1b). The same probability distribution was used for changes at all stimulus locations. The behaviourally relevant probability for the monkeys was the probability that a change would occur at a given point in time, given that no change had occurred yet (Fig. 1c). This conditional probability is called the hazard function.

Data were collected from 80 V4 neurons. Figure 2 illustrates data from a typical neuron. Responses to visually identical stimuli are aligned to stimulus onset and plotted as a function of time for two cases: when attention was directed to a stimulus within the receptive field of the cell (Fig. 2a) and when attention was directed to a stimulus outside the receptive field (Fig. 2b). In some respects, the responses in these two cases were similar—there was little spontaneous activity (grey line) before stimulus onset and a transient response to stimulus onset was followed by a moderate sustained response rate. Although the transients differed little, the sustained response was much larger when attention was directed within the receptive field of the cell (Fig. 2, compare panels a and b). An attention index was calculated as a function of time using the responses shown in Fig. 2a and b. Attentional modulation (Fig. 2c) was not maximal at time of maximal response, which occurred shortly after stimulus onset, but grew with time, paralleling the increase in the hazard function (Fig. 2d). Unexpectedly, attentional modulation was negative before stimulus onset, indicating a relative suppression of activity at the cued location. The large change in attentional modulation over the course of the trial shows that the typical measure of attentional modulation—a time-averaged mean over the entire stimulus presentation—can obscure important aspects of the modulation.

On average the attentional modulation of neurons increased gradually during trials (Fig. 3a, solid line). Although this increase over time may result from the subjects' representations of trial statistics (Fig. 3a, dotted line), it is also possible that this correlation is coincidental, and that the observed dynamics reflect a fundamental characteristic of spatial attention after stimulus onset. One reason to believe that this time course is not fundamental is the fact that other reports have suggested different modulation dynamics^{4,6,16}. To establish more conclusively the dependence of attentional modulation on task timing, we retrained the same monkeys for 3 weeks on the identical task, but with an altered hazard function (dashed line, Fig. 3b) and recorded from an additional 83 V4 neurons. Whereas the original task had an initial period of 500 ms in which there were no changes in orientation, in the modified task there was a high probability of change during this period and a period of zero probability in the subsequent 500 ms. Figure 3 illustrates the physiological effects of this change. With the modified timing, there was a sharp rise in attentional modulation shortly after stimulus onset that was not seen with the previous timing (Fig. 3, compare b and a). Moreover, in contrast to the smoothly rising modulation seen in the first recording session and in most previous reports^{4,6,7,16}, attentional modulation did not rise monotonically with time. Instead, it decreased shortly after the hazard function went to zero. For both timings, the average eye position difference between attentional states never exceeded 0.01° at any time during the trials.

Although task timing changed the time course of attentional modulation, it does not seem to be the sole determinant of the time course change. For example, in both timing sets there are pre-stimulus periods when the hazard function was zero and attentional modulation was positive, and pre-stimulus periods when the hazard function was zero and attentional modulation was negative. In schedule 1, the attentional modulation of the population of cells started to rise before the rise in the hazard function (Fig. 3a), whereas in schedule 2, the initial rise in attentional modulation

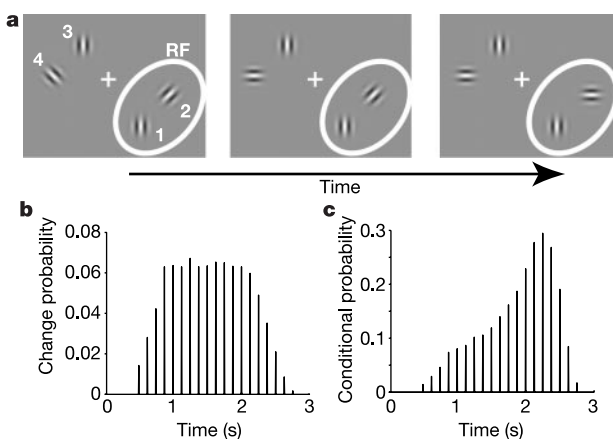


Figure 1 Orientation change detection task. **a**, Stimuli of a typical trial. Animals fixated on a central spot throughout the trial (cross). Stimuli were located at four different locations: two adjacent locations within the receptive field (oval, positions 1 and 2) and opposite the receptive field (positions 3 and 4). If the animal was cued to attend to position 2 (an 'attend in' location), it would have to release a lever when a change happened at this location (middle and right panels), while ignoring the change at position 4 (left and middle panels). Attentional modulation was measured by comparing responses for visually identical trials in which the animal attended to different locations. **b**, Orientation changes occurred at random times after stimulus onset according to the indicated probability distribution. **c**, The behaviourally relevant probability is the probability that a change will occur at a given point in time given that one has not occurred already (hazard function). The hazard function is zero for the first 500 ms (no changes occurred) and then rises to a maximum 2.25 s after stimulus presentation.

followed the rise in the hazard function by about 150 ms (Fig. 3b). To study the relationship between the hazard function and attentional modulation, correlation coefficients were computed after smoothing and shifting the two functions from the same schedule. Owing to the largely monotonic nature of both the hazard function and the attentional modulation in schedule 1, the schedule 1 functions were well correlated over a wide range of temporal shifts. On the other hand, schedule 2 correlations in both animals were maximal when the hazard function was delayed by 150 ms. It is therefore unclear whether this 150-ms delay between task timing and attentional modulation can differ between tasks, or if instead it is constant across tasks with different timings.

To quantify the relative contribution of the trained hazard functions towards attentional modulation, we applied a multiple regression model in which attentional modulation depended on the hazard functions of the two schedules. Regression coefficients were standardized to establish the relative importance of the two hazard functions in explaining attentional modulation (Table 1). For both animals and with both schedules, regression coefficients were significantly larger for the schedule type that preceded recording (Table 1, bold). This was true for a wide range of temporal offsets (100–450 ms) between the attentional modulation and the hazard functions. The attentional modulations of individual cells (Fig. 3c, d) were also correlated with the appropriate hazard function. For the first recording session, neuronal modulations were better correlated with schedule 1 than schedule 2 (above the diagonal line, Fig. 3c, paired-sample Wilcoxon $P \ll 0.01$); for the second recording session (after schedule 2 training), neuronal modulations were better correlated with schedule 2 (below the diagonal line, Fig. 3d; paired-sample Wilcoxon, $P < 0.01$). In both animals, the appropriate hazard function regression coefficient was poorer with the second schedule than the first. It is unclear whether this lower correlation is due to the more complicated temporal structure of the second schedule, or because the animals had received less training in the second schedule than in the first.

There are several reasons to doubt that this change in attentional modulation would have occurred had the animals not been exposed to schedule 2. First, the time course of the modulation after training schedule 2 is not simply a scaling or shifting of the time course seen after schedule 1. Instead, attentional modulation decreases in the

schedule 2 population of cells during an interval over which it is increasing in the schedule 1 population (750–1,000 ms). Second, if the dynamics observed after schedule 2 were the inevitable result of training, then it should have been observed in other reports. However, almost all previous reports have indicated a monotonic rise in attentional modulation after stimulus presentation. Third, if this time course was a simple consequence of training, then we might expect differences between the animals because of the different time courses of their training: animal 2 was trained with schedule 2 two months after recording the data from schedule 1, whereas animal 1 performed no tasks for almost a year in between the two schedules. Moreover, no obvious changes were observed over a period of weeks: the modulations observed during the first few weeks of recording were similar to those observed in the final weeks for both schedules. If schedule 2 attentional modulation is an inevitable result of training then we would expect the regression coefficient associated with schedule 2 to increase over time and the coefficient associated with schedule 1 to decrease over time. Instead, comparing the first and second halves of the recording sessions, the schedule 2 coefficient remained near zero in recording session 1 (−0.07 compared with 0.06) and decreased in recording session 2 (0.54 compared with 0.23), and the schedule 1 coefficient increased slightly in both sessions (session 1, 0.68 compared with 0.76; session 2, 0.06 compared with 0.16). Finally, animal 2 was trained in a different threshold level detection task in which there was a flat hazard function subsequent to these experiments, and no modulations in reaction time or attentional modulation were observed (data not shown).

To examine the specific aspects of attentional modulation that changed with the change in task timing, attentional modulation was quantified for each neuron in the two timing populations according to five different parameters (Table 2). Significant changes were seen only in those parameters related to timing. As expected, increases in

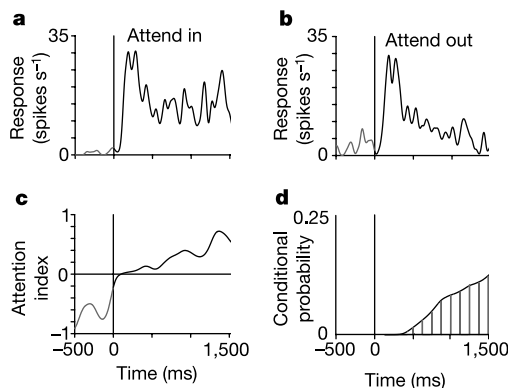


Figure 2 Dynamic attentional modulation in a V4 neuron. **a, b**, Averaged responses for trials in which a single stimulus was present in the receptive field, and a single stimulus was located in the quadrant diagonally opposite to the receptive field. For the trials included in **a**, the behaviourally relevant change occurred within the receptive field (attend in); in **b**, the change occurred in the opposite visual quadrant (attend out). **c**, Increasing effect of spatial attention with time: the attention index, based on the response difference between trials for **a** and **b**, increases during the course of a trial. **d**, The hazard function, which increases at intervals of 125 ms (unsmoothed, grey; interpolated, black) during the period over which responses were observed for this cell. All trials are aligned to stimulus onset.

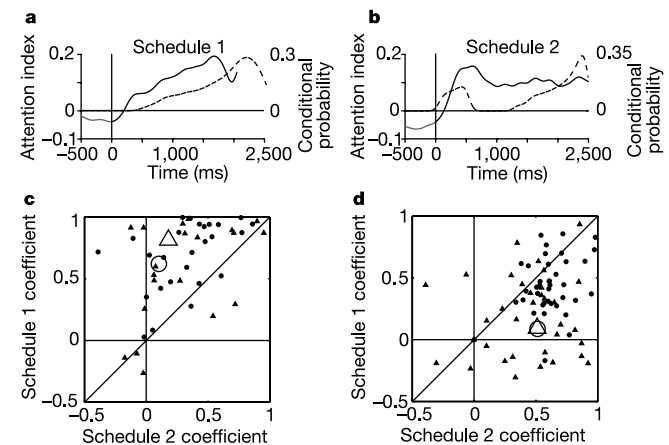


Figure 3 Attentional dynamics change according to task timing. **a, c**, Population attentional modulation averaged over all stimuli (schedule 1). **b, d**, Population attentional modulation for a sample of 83 neurons after retraining with a new task timing (schedule 2). Timing schedules (hazard functions) are shown as a dashed line (**a, b**) and attentional modulation as a solid line. All data have been smoothed by gaussian convolution ($s.d. = 75$ ms). In schedule 1, the attention index gradually rises to a maximum late in the trial (**a**). In schedule 2, the attention index rises more rapidly, reflecting an initially high change probability, and then slightly decreases during a zero change probability period starting at 500 ms. Standardized regression coefficients between the attentional modulation of individual cells and task timing show a similar change in both animals (**c, d**; circles, monkey 1; triangles, monkey 2). For the first recording session attentional modulation is significantly better correlated with schedule 1 than schedule 2 (**c**), whereas in the second recording session after schedule 2 training, the converse is true (**d**). The large symbols are the coefficients between the population responses in each animal and task timing (Table 1).

Table 1 Standardized partial regression coefficients of attentional modulation

Coefficient	Monkey 1		Monkey 2	
	Recording session 1	Recording session 2	Recording session 1	Recording session 2
Schedule 1	0.62	0.11	0.81	0.11
Schedule 2	0.10	0.52	0.18	0.51
Constant	0.25	0.33	0.01	0.32

To facilitate comparison all data were convolved with a gaussian (s.d. = 75 ms) and the hazard function was delayed to 150 ms before computing the correlation coefficient. This was the optimal delay seen in schedule 2 for both animals. Confidence intervals of 99% for all coefficients were between 0.03 and 0.04.

attentional modulation occurred significantly earlier with schedule 2 than with schedule 1. This task dependence underscores the difficulty in trying to use the latency of attentional effects as a guide to determining the physiological origin of attentional signals. On the other hand, the average magnitude of attentional modulation did not seem to vary between the two timing populations.

Previous reports have suggested, however, that attentional modulation depends on low-level factors such as the particular stimulus used to evoke responses. For example, attentional modulation can vary spatially within the receptive field⁵ and depend on the orientation¹⁷ and number of stimuli¹ within the receptive field. We studied the stimulus dependence of attentional modulation by grouping trials according to particular stimulus attributes: orientation, location and number of stimuli within the receptive field. We measured attentional modulation for single Gabor stimuli of two different orientations (preferred and null) and at two locations within the receptive field (preferred and non-preferred). To study the effect of orientation, both locations were considered, whereas to study the effect of location, both orientations were considered. The mean response difference between preferred and null orientation was 33%; between preferred and non-preferred location it was 20%. Both the time course and magnitude of attentional modulation were similar for preferred and non-preferred orientations (Fig. 4a, d). This is consistent with the finding that the magnitude of V4 attentional modulation is independent of the stimulus orientation⁶. Similarly, both the magnitude and the time course of attentional modulation were well matched at the preferred and non-preferred locations (Fig. 4b, e). Finally, the attentional modulation seen with two Gabor stimuli within the receptive field was similar in time course and magnitude to the attentional modulation seen when only one Gabor stimulus was present in the receptive field (Fig. 4c). Although contrary to some reports^{1,4}, these data are consistent with other reports showing significant attentional modulation when only a single stimulus is present within the receptive field^{2,6,7,16,18}. In all of these cases, the relative contribution of stimulus variation, as measured by standardized regression coefficients, was significantly smaller than the contribution of the appropriate hazard function to attentional modulation. These results show that the attentional modulation of a neuronal population can be relatively unaffected by many variations in stimulus configuration, including the number of stimuli within a receptive field. Moreover, they demonstrate that such constancy can hold even when retraining causes large changes in the dynamics of attentional modulation. One unresolved issue is whether such consistency would be visible throughout visual cortex, even among neuronal signals not used in the task or with very different maximal attentional modulations.

One observation that has not been reported previously is that the average attentional modulation at a behaviourally relevant location can be negative for periods of time. This relative suppression of spontaneous activity at the attended location was seen across both timing sets and all stimuli (Fig. 4). Because the attentional modulation metric involves a subtraction between responses in which spatial attention has been shifted, it is unclear whether this negative value reflects a transient facilitation of responses at unattended locations, or a suppression of responses at the attended location. For example, suppression of visual responses to stimuli presented at previously attended locations has been observed in area 7a^{19,20}. Whether it reflects spatially specific facilitation or suppression, the cause of the negative modulation is unclear. It is not a direct effect of cueing²¹, because a cue stimulus was not presented at the beginning of the trials.

Our results show that subjects form an internal representation of task timing acquired during the course of training, and then use this representation to temporally modulate behaviourally relevant visual responses. Unlike the spatial cues, there was no explicit requirement for the animals to monitor task timing and modulate their attention accordingly. Moreover, task timing was changed without any explicit cue or instruction to alert the animals to the change. The consistency of results between the animals therefore suggests that temporal strategies are commonly adopted, and that these strategies reflect task timing. Thus, just as task difficulty can affect the magnitude of attentional modulations¹⁷, so can task timing affect the timing of such modulations. Although the two monkeys were able to perform the required task at an accuracy level of 90% after only about an hour of training on the second timing, it remains to be established whether attentional modulations can change on a similar timescale.

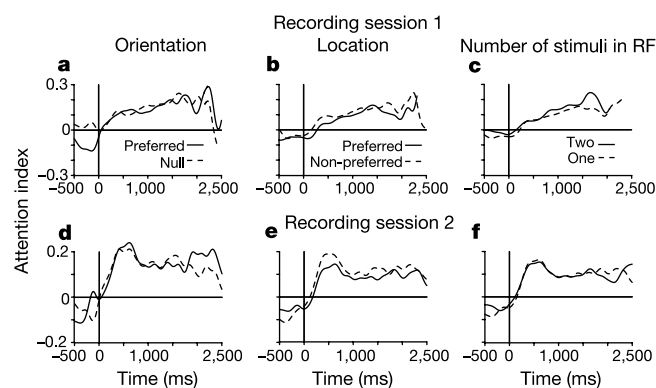


Figure 4 Population attentional modulation with different stimuli. **a–c**, Average attention index for V4 neurons from two animals after schedule 1 training. **d–f**, Index for the population of neurons studied after schedule 2 training. In **b** and **e**, modulation is examined according to whether or not the attended stimulus was at the preferred location within the receptive field; in **a** and **d**, according to the orientation of the stimuli; and in **c** and **f**, according to stimulus number within the receptive field (RF). Both the time course and the magnitude of attentional modulation in the neuronal population are similar for different stimuli (solid versus dashed lines). For all stimuli, the time course of attentional modulation is affected similarly by the change in task timing.

Table 2 Median parameters of attentional modulation index among single neurons

Parameter	Recording session 1	Recording session 2
Time of zero crossing (ms)*	225	168
Time of half-maximum (ms)†	415	254
Magnitude during pre-stimulus period	–0.06	–0.07
Maximum magnitude	0.23	0.22
Slope at zero crossing (s ^{–1})	0.63	0.72

*Wilcoxon, $P < 0.02$.

†Wilcoxon, $P < 0.001$.

Whatever their origins, the existence of rapid response modulations shows that the effects of attention on individual cells or neuronal populations cannot be fully characterized by a single constant number that applies to all tasks. Such modulations pose a challenge to interpreting physiological data when behavioural timing is not strongly constrained. For example, when stimuli are presented for hundreds of milliseconds, the animal might attend only within a particular period of time such as the beginning or end of the presentation, or alternatively might randomly choose different periods from trial to trial in which to concentrate perceptual resources. When animals are free to use different temporal strategies, the neuronal modulations observed from trial to trial, or from cell to cell, might reflect a choice of strategy more than a fundamental characteristic of spatial attention. Such task-related changes in strategy can have a much larger effect on the attentional modulation of sensory responses than changes in sensory input itself (Fig. 4). Given that temporal strategies will probably arise in almost any task with which a subject becomes familiar, such effects should be visible across a broad range of sensory modalities and tasks. Our results highlight that extraretinal effects on sensory signals reflect not only immediately defined task requirements such as the location or content of a cue, but also strategies based on accumulated experience. □

Methods

Visual stimuli

The stimuli were Gabor stimuli of sinusoidally varying contrast (4 Hz, 100% peak contrast) truncated at a radius of twice the standard deviation of the gaussian envelope of the Gabor. Gabor stimuli were modulated around a mean luminance achromatic point in colour space that also defined the background. Single Gabor stimuli of varying chromatic modulation, size, spatial frequency, location and orientation were used to characterize receptive fields and specify the stimulus parameters used in the attention task. Spatial receptive fields were mapped quantitatively using single Gabor stimuli. We tested all well-isolated, visually responsive cells for attentional effects.

Task design

Two monkeys (*Macaca mulatta*) performed an orientation change detection task. Trials were presented in block mode (12 or 15 trials), in which the behaviourally relevant location was fixed within each block. Animals were required to fixate on a small dot ($\sim 0.1^\circ$) throughout each trial (fixation window width = 1.0° to 1.4°). Attention was spatially cued using instruction trials at the beginning of each block, in which only a single stimulus was presented. Subsequent trials within the block, which were used in the data presented here, included stimuli at this cued location as well as at other locations. Stimuli were presented at two non-overlapping locations within the receptive field. These adjacent locations were always at the same eccentricity (between 2° and 5°) and offset symmetrically from the receptive field centre (positions 1 and 2 in Fig. 1a). In addition to these receptive field stimuli, stimuli were presented simultaneously at symmetrical locations in the quadrant diagonally opposite from the receptive field, so that during each trial there were either two or four Gabor stimuli present. Trials with two Gabor stimuli were randomly interleaved with trials in which four Gabor stimuli were presented. Cue location randomly switched between the four possible locations on block completion. Cells used in this study had at least eight blocks completed for each cued location.

The monkeys' task was to release a lever as soon as a change occurred at the cued location, while ignoring changes occurring at any other location. Orientation changes occurred only when the counterphasing Gabor stimuli reached zero contrast (every 125 ms, see Fig. 1c, d) and changes at different locations were forced to occur at different times. At each location no more than one orientation change occurred during the trial. Animals were rewarded with juice when they released a lever between 250 and 450 ms after a change at the cued location. Releases at any other time or eye movements outside the fixation window immediately terminated the trial without reward. About 10% of trials were 'catch' trials in which no change occurred at the cued location and the monkey was rewarded for keeping the lever depressed and maintaining fixation. The monkeys' performance, excluding fixation breaks, was greater than 90% for all data presented here and was independent of the time at which the behaviourally relevant change occurred.

Recording

Recordings were made from individual neurons in area V4 on the prelunate gyrus in daily sessions using transdural electrodes (0.5–1.5 M Ω at 1 kHz) and conventional extracellular recording techniques. Action potentials were recorded with a resolution of 1 ms, synchronized with the vertical retrace of the monitor. Eye position was monitored by scleral search coil. Eye position and lever releases were recorded with a resolution of 5 ms.

Data analysis

Attentional modulation was measured by comparing responses to identical stimulus configurations in different cue-location blocks. Modulation was quantified using an attention index $A(t) = (R_A(t) - R_B(t)) / (R_A(t) + R_B(t))$, where R_A and R_B were the

responses when attention was directed to locations A (within the receptive field) and B (opposite the receptive field), respectively. To avoid the attention index being dominated by optimal stimulus combinations, attentional modulation was computed for each stimulus combination separately before averaging. For most cells, the compared trials also had identical change times for all stimuli. Because trials were of variable length, averages were computed according to the number of trials contributing to each time point. All data shown are based on trials truncated at the first change that occurred within the receptive field, whether or not that change was at the cued location. The relevant probability within a change detection trials is the conditional probability that a change will occur at a given time in the trial given that a change has not already occurred. This is called the hazard function and is given by $H(t) = p(t) / (1 - \int_0^t p(x) dx)$, where $p(t)$ is the change probability function. The hazard functions of this study go to zero because long trials are likely to be catch trials in which no orientation change takes place at the cued location. Because of the catch trials the total probability of change occurring within a trial is less than one. Correlation coefficients between task timing and physiological measurements were computed for a range of gaussian filters of varying standard deviation (1, 10, 25, 50, 75, 100, 150, 200 ms). Correlations were low for filter sizes for less than 50 ms across all temporal offsets but were about 0.7 for the optimal offset at a filter width of 75 ms and continued to rise for wider filters. For Figures 3 and 4 we used a filter width of 75 ms, which is similar to the fastest timescale over which spatial shifts in attention can be observed behaviourally^{22–24}. To quantify the relationship between attentional modulation and the hazard functions, we computed multiple regression coefficients on the basis of a model in which attentional modulation was modelled as a linear sum of two hazard functions plus a constant: $A(t) = b_1 H_1(t - \tau) + b_2 H_2(t - \tau) + b_3$. Attentional modulation and hazard functions were convolved with gaussian filters of 75 ms s.d. and shifted before regression. The optimal temporal shift τ of the hazard functions was evaluated by minimizing root mean square (r.m.s.) error as a function of τ and the b coefficients using the downhill simplex method. The maximum allowable shift was 150 ms in accordance with the optimal schedule 2 correlations. To evaluate the relative importance of the two coefficients (Fig. 3 and Table 1), they were normalized according to the standard deviations (σ) over time of the hazard and attentional modulation functions to compute standardized partial regression coefficients ($\beta_i = b_i \sigma_{H_i} / \sigma_A$). Stimulus effects were evaluated by incorporating additional terms into the attentional modulation model: $A(t) = b_1 H_1(t) + b_2 H_2(t) + b_3 H_1(t)S + b_4 H_2(t)S + b_5 S + b_6$ where S was a dummy variable coding for the examined stimulus variation (orientation, location or stimulus number) with values of zero and one.

Received 22 April; accepted 5 August 2002; doi:10.1038/nature01057.

- Moran, J. & Desimone, R. Selective attention gates visual processing in the extrastriate cortex. *Science* **229**, 782–784 (1985).
- Motter, B. Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. *J. Neurophysiol.* **70**, 909–919 (1993).
- Desimone, R. & Duncan, J. Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* **18**, 193–222 (1995).
- Luck, S., Chelazzi, L., Hillyard, S. & Desimone, R. Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *J. Neurophysiol.* **77**, 24–42 (1997).
- Connor, C., Preddie, D., Gallant, J. & Van Essen, D. Spatial attention effects in macaque area V4. *J. Neurosci.* **17**, 3201–3214 (1997).
- McAdams, C. & Maunsell, J. Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *J. Neurosci.* **19**, 431–441 (1999).
- True, S. & Maunsell, J. Effects of attention on the processing of motion in macaque middle temporal and medial superior temporal visual cortical areas. *J. Neurosci.* **19**, 7591–7602 (1999).
- Colby, C. & Goldberg, M. Space and attention in parietal cortex. *Annu. Rev. Neurosci.* **22**, 319–349 (1999).
- Coull, J. T., Frith, C. D., Buchel, C. & Nobre, A. C. Orienting attention in time: behavioural and neuroanatomical distinction between exogenous and endogenous shifts. *Neuropsychologia* **38**, 808–819 (2000).
- Nobre, A. C. Orienting attention to instants in time. *Neuropsychologia* **39**, 1317–1328 (2001).
- Treisman, M., Faulkner, A., Naish, P. & Brogan, D. The internal clock: evidence for a temporal oscillator underlying time perception with some estimates of its characteristic frequency. *Perception* **19**, 705–743 (1990).
- Zakay, D. & Block, R. The role of attention in time estimation processes in *Time, Internal Clocks and Movement* (eds Pastor, M. A. & Artieda, J.) 143–164 (Elsevier Science, New York, 1996).
- Harrington, D., Haaland, K. & Knight, R. Cortical networks underlying mechanisms of time perception. *J. Neurosci.* **18**, 1085–1095 (1998).
- Westheimer, G. Visual signals used in time-interval discrimination. *Vis. Neurosci.* **17**, 551–556 (2000).
- Migliore, M., Messineo, L., Cardaci, M. & Ayala, G. Quantitative modeling of perception and production of time intervals. *J. Neurophysiol.* **86**, 2754–2760 (2001).
- Seidemann, E. & Newsome, W. Effect of spatial attention on the responses of area MT neurons. *J. Neurophysiol.* **81**, 1783–1794 (1999).
- Spitzer, H., Desimone, R. & Moran, J. Increased attention enhances both behavioural and neuronal performance. *Science* **240**, 338–340 (1988).
- Colby, C., Duhamel, J. & Goldberg, M. Visual, presaccadic, and cognitive activation of single neurons in monkey lateral intraparietal area. *J. Neurophysiol.* **76**, 2841–2852 (1996).
- Steinmetz, M., Connor, C., Constantinidis, C. & McLaughlin, J. Covert attention suppresses neuronal responses in area 7a of the posterior parietal cortex. *J. Neurophysiol.* **72**, 1020–1023 (1994).
- Constantinidis, C. & Steinmetz, M. Neuronal responses in area 7a to multiple stimulus displays: ii. responses are suppressed at the cued location. *Cereb. Cortex* **11**, 592–597 (2001).
- Robinson, D., Bowman, E. & Kertzman, C. Covert orienting of attention in macaques. II. Contributions of parietal cortex. *J. Neurophysiol.* **74**, 698–712 (1995).
- Krose, B. & Julesz, B. The control and speed of shifts of attention. *Vision Res.* **29**, 1607–1619 (1989).
- Saarienen, J. & Julesz, B. The speed of attentional shifts in the visual field. *Proc. Natl Acad. Sci. USA* **88**, 1812–1814 (1991).
- Verstraten, F., Cavanagh, P. & Labianca, A. Limits of attentive tracking reveal temporal properties of attention. *Vision Res.* **40**, 3651–3664 (2000).

Acknowledgements This work was supported by grants from the National Eye Institute and the Human Frontier Science Program. J.H.R.M. is an Investigator with the Howard Hughes Medical Institute. We thank D. Murray and T. Williford for assistance with the animals, and W. Bosking, E. Cook, R. A. Eatock, M. Shadlen, D. Sparks, T. Yang and T. Williford for comments on the manuscript.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.M.G. (e-mail: gghose@bcm.tmc.edu)

FGFR-related gene *nou-darake* restricts brain tissues to the head region of planarians

Francesc Cebrià^{*†}, Chiyo Kobayashi^{*†}, Yoshihiko Umesono^{*}, Masumi Nakazawa[‡], Katsuhiko Mineta[§], Kazuho Ikee[§], Takashi Gojobori[§], Mari Itoh^{||}, Masanori Taira^{||}, Alejandro Sánchez Alvarado[¶] & Kiyokazu Agata^{*#}

^{*} Group for Evolutionary Regeneration Biology, Center for Developmental Biology RIKEN Kobe, 2-2-3 Minatogima-minamimachi, Kobe 650-0047, Japan

[‡] Center for Information Biology and DNA Data Bank of Japan, National Institute of Genetics, and [§] Department of Genetics, The Graduate University for Advanced Studies, Yata 1111, Mishima 411-8540, Japan

^{||} Department of Biological Sciences, Graduate School of Science, University of Tokyo, and Core Research for Evolutional Science and Technology (CREST), Japan Science and Technology Corporation, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

[¶] Department of Neurobiology and Anatomy, University of Utah School of Medicine, 50 N. Medical Drive, Salt Lake City, Utah 84132, USA

[#] Department of Biology, Faculty of Science, Okayama University, 3-1-1 Tsushima-naka, Okayama 700-8530, Japan

[†] These authors contributed equally to this work

The study of planarian regeneration may help us to understand how we can rebuild organs and tissues after injury, disease or ageing¹. The robust regenerative abilities of planarians are based upon a population of totipotent stem cells (neoblasts)^{2–4}, and among the organs regenerated by these animals is a well-organized central nervous system^{5,6}. In recent years, methodologies such as whole-mount *in situ* hybridizations and double-stranded RNA have been extended to planarians with the aim of unravelling the molecular basis of their regenerative capacities^{7–11}. Here we report the identification and characterization of *nou-darake* (*ndk*), a gene encoding a fibroblast growth factor receptor (FGFR)-like molecule specifically expressed in the head region of the planarian *Dugesia japonica*. Loss of function of *ndk* by RNA interference results in the induction of ectopic brain tissues throughout the body. This ectopic brain formation was suppressed by inhibition of two planarian FGFR homologues (*FGFR1* and *FGFR2*). Additionally, *ndk* inhibits FGF signalling in *Xenopus* embryos. The data suggest that *ndk* may modulate FGF signalling in stem cells to restrict brain tissues to the head region of planarians.

Planarians have a well-organized and molecularly complex^{12,13} central nervous system (CNS), which, in *D. japonica*, consists of two lobes that connect at their most anterior ends to form an inverted U-shaped brain. Each lobe is connected to a ventral nerve cord that traverses the animal along its anterior–posterior axis, and 9 lateral branches project away from each lobe towards the periphery of the head⁵. In order to identify and characterize molecules involved

in the process of brain regeneration in planarians, we prepared microarrays containing 1,640 non-redundant transcripts derived from a head complementary DNA library (M.N. and K.M., unpublished results). Here we present the characterization of one such gene, which we have named “*nou-darake*” (“brains everywhere”, in Japanese); the phenotype was observed in specimens of *D. japonica* that had been injected with double-stranded RNA (dsRNA).

The gene *ndk* is highly and specifically expressed in the head in both brain and non-brain tissues (Fig. 1a). During regeneration, *ndk* expression is first detected 24 h after amputation only in anterior blastemal cells, including the new brain primordium⁶. Sequence analyses reveal that *ndk* codes for a putative transmembrane protein with two extracellular immunoglobulin-like domains related to FGF receptors (Figs 1b and 2), but significantly divergent from the IgG domains found in two planarian FGFR-like receptor homologues recently isolated¹⁴. But NDK lacks the cytoplasmic kinase domains characteristic of this receptor family (Fig. 1b). Instead, the cytoplasmic domain of NDK is short, rich in serine residues, and has no significant homology to known sequences. When the entire deduced amino-acid sequence of *ndk* is considered, the highest similarity is found to the human *FGF receptor-like 1* (*FGFRL1*) gene¹⁵.

We analysed the function of *ndk* in planarians by RNA interference (RNAi)^{10,16}. Animals were injected with *ndk* dsRNA and subsequently amputated (see Methods). Silencing of *ndk* expression in dsRNA-injected animals was confirmed by whole-mount *in situ* hybridizations (Fig. 3a–d). Seven days after amputation, ectopic eyes began to differentiate in dorso-posterior regions of the body. After 15 d of regeneration, 94% of the injected animals (108/115) formed ectopic eyes (Fig. 3h), consisting of both pigment (Fig. 3h, arrows) and visual cells with projecting axons (Fig. 3i, j, arrowheads). In all of the injected animals (115/115), we also found ectopic brain tissues differentiating in posterior regions of the planarian body (Fig. 3j, l, n, p; arrows). Ectopic brain cells express specific genes of both lateral branches (Fig. 3n) and the central region (Fig. 3p) of the brain. In trunk pieces allowed to regenerate

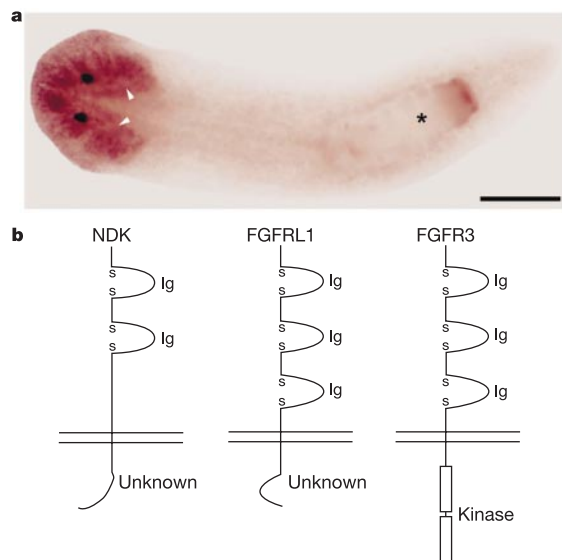


Figure 1 Expression pattern of *ndk* and predicted structure. **a**, *ndk* is expressed in the head region in both the inverted U-shaped brain (arrowheads) and non-brain cells. Signal detected in the pharynx (asterisk) is probably a non-specific signal due to trapped probe. Anterior to the left. Scale bar, 1 mm. **b**, Two extracellular IgG domains are predicted in NDK.

bipolarly, we observed both a posterior expansion of the regenerated brain tissues, and the *de novo* appearance of brain cells in areas near or within the posterior blastema (Fig. 3l, n).

In animals injected with *ndk* dsRNA, the early stages of regeneration are normal and new brain tissues are restricted to the blastema; at day 4–5 of regeneration, however, brain-like structures differentiate outside the blastema, and as regeneration proceeds the ectopic brain gradually expands into more posterior regions (Supplementary Fig. 1). The new pharynx appears normally in the new central region in regenerating head and tail fragments, suggesting that the ectopic appearance of brain cells in posterior regions does not disturb the relative proportions of the planarian body (Supplementary Fig. 1). Differentiation of ectopic brain tissues and extra eyes were also observed in intact, non-regenerating animals after injection of *ndk* dsRNA (Fig. 4b, d, f). To test whether only the brain or the whole anterior region is expanded, we analysed the expression of clone 0821_HN (unknown gene), which is specifically expressed in the head periphery (Fig. 4e). After *ndk* dsRNA injection, the expression pattern of this gene is not changed (Fig. 4f), suggesting that the head region is not expanded. These results suggest that ectopic brain formation can occur without disturbing head–trunk identities. Planarian neoblasts can be specifically eliminated by X-ray irradiation, resulting in the death of the animals after a few weeks^{11,14}. In irradiated animals that had been injected with *ndk* dsRNA, ectopic brain formation was completely absent or significantly reduced (data not shown), clearly implicating neoblasts in ectopic brain formation.

Because of the structural similarities between NDK and FGFRs, we wondered whether *ndk* might be modulating a putative planarian FGF pathway. The recent identification of two FGFR homologues in *D. japonica* and their expression in both brain and neoblasts indicate that FGF-like signalling pathways may be operating in planarians¹⁴. The function of these genes, however, remains unknown, as no detectable abnormality was found after RNAi experiments¹⁴. In double *ndk/FGFR1* and *ndk/FGFR2* dsRNA-injected animals, ectopic brain tissues could be detected, although at a level slightly lower when compared to *ndk*-dsRNA-injected

NDK	1	WNVFISWMLISYKSKHCTIEVVKPKPDITNSGSELK-----ICPT
FGFR1	1	--MTPSEILILLLPPLILGAPPAARAGPPKADKVV-----ROV
FGFR3	1	VGAPACALHICVAIAVAGASSESLTEQRVVGRAAEVSGPEPGQCPQV
NDK	46	VLSREGDANVFNSLMYEWNTKGLTEVSVQVNPKISHTENNRLKLMKEV
FGFR1	41	ARIGRTVRLQCEVGGDPPPLTMMTKDGRTHSGNSRERVLPQGLAVKOV
FGFR3	51	FCSGDAVELSCPPGCGGPGMPTVWVKDC-TGLVPSERVIVSGPQRLVDNA
NDK	96	TSDESGVYCKCVTFGFSREVVENVETIDSLQERLCARKNPKDSKLDPPC
FGFR1	90	BREDAGVYCKATNFGSTISVNYTIVVLDITSEKESLCHDSSSGQEDP
FGFR3	100	SHEDSGVYCROR-LTQVRLCHFVVRVDAPSSGDDPDGDAEDTGVDT
NDK	146	FVDPQVRS-----GKITIRKRVGSTVRLCDATGSPOLKYLWLGKT
FGFR1	140	ASQQWAPRPTQPSKMRVRVIARPVGSSVRLKCVASGHPRPITIMRDDQ
FGFR3	149	GAPYWRP-----BRMKKLLAVPAANTVRERCPAAGNTPPSISWLNGR
NDK	188	IAN---WQNVGTEGPVHTIRNLKSHYTGQYTCOVENKAGSNNTYOL
FGFR1	190	ALT---RPEAEPKPKFWLISLKNLRPEDSKRYTCRVSNRAGINAPYKV
FGFR3	194	EFRGEHRIGGKIKLRHQQWLSLVESVPSDRGNYTCVVENKESIRQYTL
NDK	234	IVDPSNNSPKLTDLSNYTFRSKENISISVIEICECKEPIIRWLKRVEN
FGFR1	237	DVIGRTRSKPVLTCTHEVNTVDSGTTISFOCKYRSVDVGVQLWLRVVE
FGFR3	244	DVLEERSPHRPILQAGLPANTAVLGSVDVDFHCKVYSAPQFIQWLKEVE
NDK	284	FDSSNKEQSNGLVLPNAKDSHCELYVLDTCVSTRIGQPNLEKTLINME
FGFR1	287	GEGR-----HNSITDGGQKEVWLETCGVWSRPGGSSNKLITR
FGFR3	294	NCS-----KVGPDCTPYVTVLKT--AGANTTDKELEVLISHN
NDK	334	SVETQCSGKYVVMALNNTIVNLNFEHVVYITIEPDKDAIGI-----
FGFR1	328	-ARQDDAGNYTCLGANMGYSRSAFTIVLPDPKPGGPPVSSSSATSLP
FGFR3	329	-VTEPDAGNYTCLAGNIGESHSANLVVLPAAEELVLEADEAGSVYAG--
NDK	--	--
FGFR1	377	WP
FGFR3	--	--

Figure 2 Sequence comparison between the extracellular domains of NDK and the proteins encoded by human *FGFR1* and *FGFR3* genes. Identical or similar amino-acid residues are indicated in black or grey, respectively.

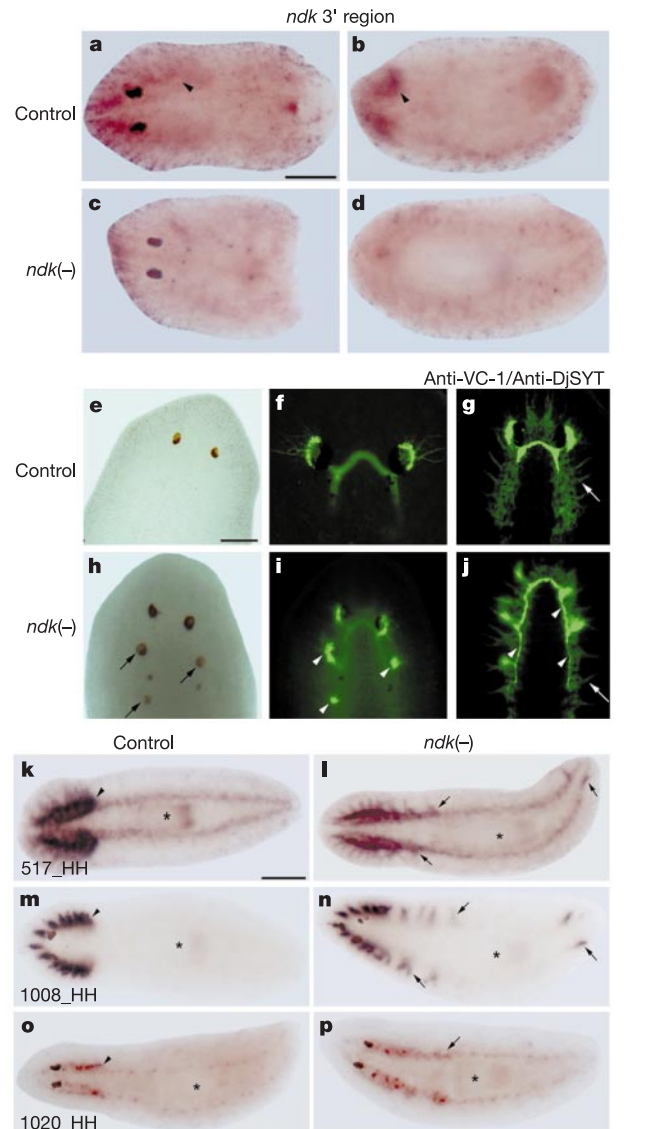


Figure 3 Effects of *ndk*-dsRNA injections during regeneration. **a–d**, Such injections silence the *ndk* endogenous expression. In control animals, *ndk* is expressed in the brain (arrowheads) of both head (**a**) and trunk (**b**) pieces after 3 days of regeneration. No expression, in contrast, is detected in *ndk*-dsRNA-injected head (**c**) and trunk (**d**) regenerants. **e–j**, Ectopic eyes appear in injected animals. Bright field view of control (**e**) and *ndk*-dsRNA-injected animals (**h**) after 16 days of regeneration. In dsRNA-injected animals, ectopic eyes differentiate in posterior regions. (**h**, arrows). Double whole-mount immunostaining with anti-DjSYT²⁷ and anti-VC-1 (ref. 28) antibodies on control (**f**, **g**) and injected animals (**i**, **j**). Ectopic visual cells detected by anti-VC-1 are dorsally located (arrowheads in **i**), and project visual axons to more ventral positions (arrowheads in **j**). Animals injected with *ndk*-dsRNA also show an abnormal regenerated brain after immunostaining with anti-DjSYT. In control animals, a normal brain is regenerated (**g**, pale green; arrow points to lateral branches). In injected animals, brain-like tissues appear more posterior (arrow in **j**). Ectopic brain is detected for different neural-specific markers in animals that have been regenerating for 15 days. In control animals, trunk pieces regenerate a normal brain, as shown after *in situ* hybridization for 517HH (**k**, protein tyrosine phosphatase receptor homologue and pan-neurally expressed), 1008HH (**m**, glutamate receptor homologue and expressed in brain lateral branches), and 1020HH (**o**, unknown and expressed in the central spongy region of the brain). Arrowheads in **k**, **m** and **o** point to the posterior end of the regenerated brain. In *ndk*-dsRNA-injected animals, unambiguous brain expansion is detected for the same 3 neural markers (**l**, **n**, **p**, respectively). Arrows in **l**, **n** and **p** point to ectopic brain structures. In **k–p**, asterisks mark the pharynx. In **a–d** and **k–p**, anterior, to the left; in **e–j**, anterior is to the top. Scale bars: **a–d**, 0.3 mm; **e** and **h**, 0.4 mm; and **k–p**, 1 mm.

animals (Fig. 5). In triple-injected animals, however, no ectopic brain tissues were detected in posterior body regions in any of the regenerating trunk and tail fragments (17/17) (Fig. 5h). In 5/5 of triple-injected regenerating head fragments, some level of ectopic brain tissues could be detected (Fig. 5g). These results suggest that *ndk* negatively regulates the FGFR signalling in the trunk, which are required for the ectopic brain formation.

Because it is not yet possible to perform gain-of-function analyses in planarians, we carried out *ndk* messenger RNA injection experiments in *Xenopus* embryos in order to test whether or not *ndk* could modulate FGF signalling pathways, as well as to analyse the degree of functional conservation on FGF signalling. Injection of *ndk* mRNA in dorsal blastomeres arrested gastrulation movements, leading to open yolk plug phenotypes (data not shown) that are indistinguishable from those obtained by the introduction of the dominant-negative FGF receptor (XFD)¹⁷. Examination of the expression pattern of the early mesoderm marker *brachyury* (*Xenopus bra*) uncovered a suppression of *bra* expression (blue) by NDK as well as XFD in the region where the nuclear β -galactosidase lineage tracer is present (visualized in red; Fig. 6a). We also found that the effects of *ndk* are still observed when both its intracellular and transmembrane regions were deleted (Fig. 6a), indicating that the extracellular domain of NDK retains the inhibitory activity against the *bra* expression¹⁸.

As the structural features of NDK suggest that this protein may be interacting with FGF signalling, and FGF is known to induce *bra* expression in *Xenopus* animal cap explants¹⁹, we next examined whether NDK could inhibit *bra* expression by FGF. Basic FGF (bFGF) treatment induces *bra* expression in animal caps, and this expression is inhibited by *ndk* in a fashion that is similar to XFD²⁰ (Fig. 6b). In addition, we have so far been unable to co-immunoprecipitate Flag-tagged eFGF after expression of myc-tagged NDK. These results do not exclude the possibility that NDK might modulate FGF signalling by a regulatory mechanism involving the interaction of signalling receptors with a pseudoreceptor, in a similar fashion to the inhibition of TGF- β signalling by the pseudoreceptor BAMBI²¹. Recently, an inhibitory molecule of FGF signalling, named Sef, has been identified^{22,23}. Whereas its intracellular domain is required for inhibiting FGF signalling, NDK can inhibit FGF signalling when lacking its intracellular domain (Fig. 6a). Thus, *ndk* may be a new modulator of the FGF signalling pathway.

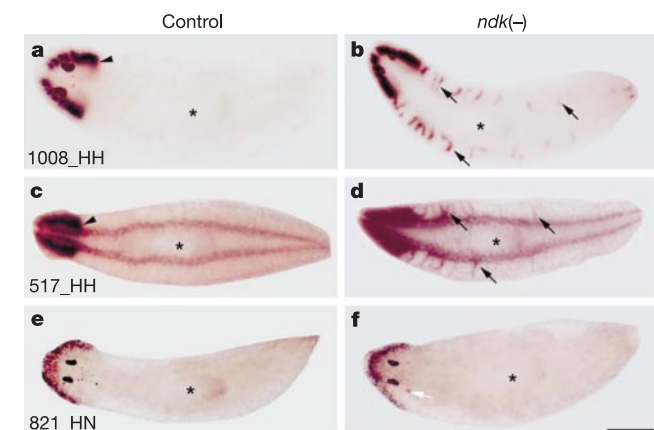


Figure 4 Brain expansion in intact animals. Ten days post-injection, ectopic brain structures (arrows in **b, d**) are detected by *in situ* hybridization with clones 1008_HH (**b**) and 517_HH (**d**). Arrowheads in **a–c** indicate the posterior border of the brain in control animals. Clone 821_HH is expressed in the head periphery of intact and control animals (**e**). This head marker is not expanded posteriorly in *ndk*-dsRNA-injected animals (**f**), even though they show ectopic eyes (white arrow in **f**), which suggest that brain expansion has taken place in them. In all figures, anterior is to the left. Scale bar, 1 mm.

How could the silencing of a gene specifically expressed in the planarian head lead to the differentiation of brain-like tissues throughout the body in a non-autonomous cell manner? Sequence analyses, *in situ* hybridization data, RNAi experiments and mRNA injections in *Xenopus* show that NDK has potential FGF binding domains (Fig. 2), is expressed in the head of planarians (Fig. 1a), restricts brain differentiation to the planarian head region (Figs 3 and 4), and is capable of inhibiting FGF signalling (Fig. 6). Then, one simple model would be that NDK may restrict the diffusion range from a putative source of brain-inducing factors (FGF or FGF-like molecules) in the planarian head to the rest of the body. Loss of function of *ndk*, therefore, would allow these factors to diffuse to posterior regions, and interact with FGF receptors outside of the head region, thus triggering ectopic brain formation. Our observation of a gradual brain expansion to more posterior regions in dsRNA-injected animals supports this idea. As these hypothetical brain-inducing factors must diffuse distances of several millimetres between the planarian head and the posterior regions where the ectopic brain is formed, diffusion rates as well as the role of extracellular matrix components²⁴ during this process should be analysed.

Even though we have yet to identify FGF-like ligands in planarians, *ndk* provides strong molecular evidence for the existence of a brain-inducing circuit based on the FGF signalling pathway. Whereas antagonists of BMP4 are believed to be the main neural inducers in vertebrates, recent work has suggested that FGFs are important in neural formation and patterning. For instance, BMP antagonists do not induce neural tissues in the presence of dominant-negative FGF receptors in *Xenopus*²⁵. In addition, studies in chicken have shown that neural induction by BMP antagonists requires FGF signalling²⁶, suggesting that FGFs, as in planarians, may be essential for neural tissue formation in the vertebrates. The

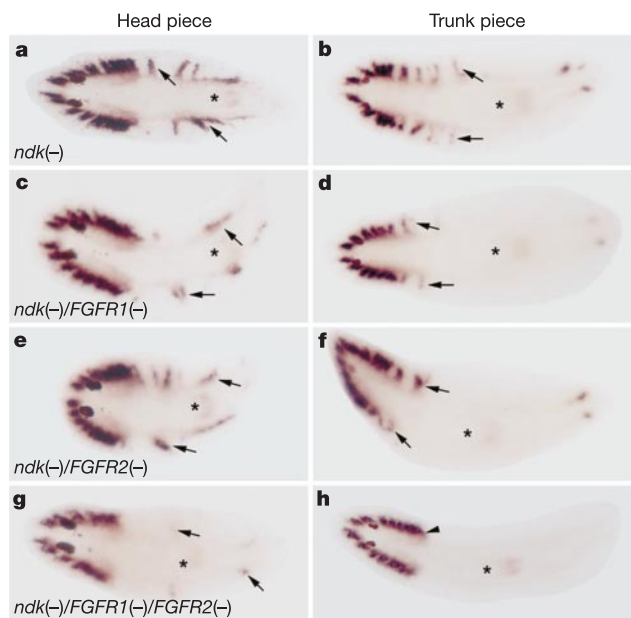


Figure 5 Effects of *ndk*- and *FGFR*-dsRNA injections on regenerating head (**a, c, e, g**) and trunk (**b, d, f, h**) fragments. In *ndk*/*FGFR1* (**c, d**) and *ndk*/*FGFR2* (**e, f**) double dsRNA-injected samples ectopic brain tissues appear in posterior regions, although the levels of ectopic brain in the trunks appear slightly lower when compared to *ndk*-dsRNA-injected animals (**b**). In *ndk*/*FGFR1*/*FGFR2* (**g, h**) triple dsRNA-injected animals, no ectopic brain tissues are detected in regenerating trunks (**h**), but low level of brain expansion occurs in regenerating heads (**g**). All samples correspond to 15-day regenerants. Arrows indicate ectopic brain tissues. Arrowhead in **h** indicates the posterior border of the brain. Asterisks mark the pharynx. Anterior to the left.

fact that planarian *ndk* can functionally inhibit *Xenopus bra* activation during *Xenopus* gastrulation raises the possibility that the vertebrate homologue of *ndk* may be modulating FGF signalling. Further studies will be required to understand the extent of the involvement of *ndk* in vertebrate organogenesis—in particular, neurogenesis. □

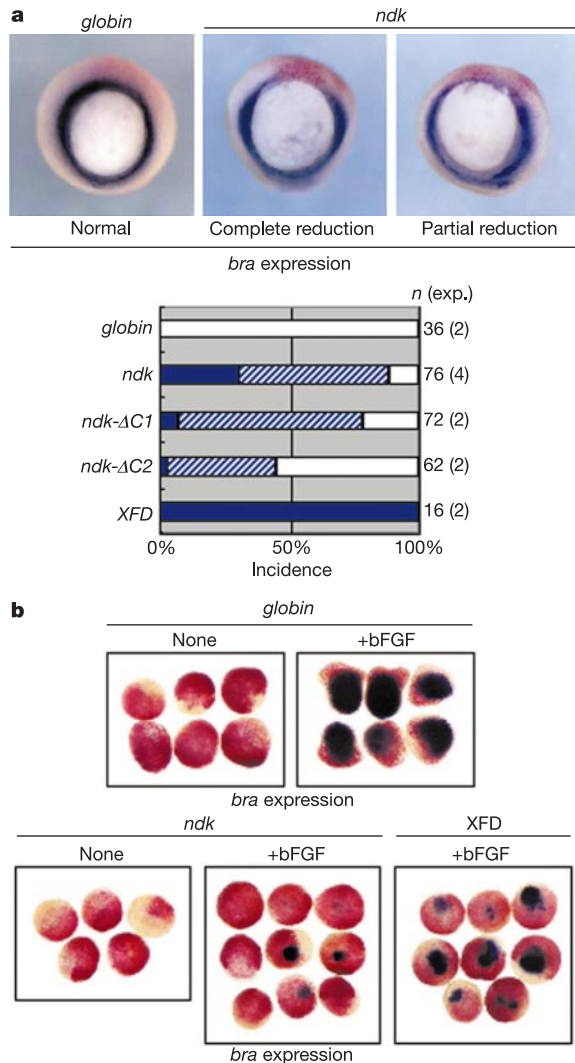


Figure 6 Effects of *ndk* mRNA injections into *Xenopus* embryos. **a**, *ndk* inhibits *Xenopus bra* expression in whole embryos. Embryos were injected with mRNA for *ndk* and nuclear β -galactosidase (β -gal) as a tracer into the dorsal marginal zone at the 4-cell stage, fixed at the gastrula stage (stage 11), stained with Red-gal as a substrate for β -gal, and analysed by whole-mount *in situ* hybridization for *bra*. Levels of *bra* expression in the region where β -gal was stained in red are categorized into normal expression (open bar) and complete (filled bar) or partial (striped bar) reduction as indicated in the middle three bars. *n*, number of embryos analysed; (exp.), number of experiments. Note that deletion of both intracellular domain (*ndk*-ΔC1) and transmembrane domain (*ndk*-ΔC2) reduces the activity. Amounts of mRNA (pg per embryo): *globin* (negative control) and *ndk* constructs, 500–1,000. **b**, *ndk* inhibits *bra* expression in animal caps treated with bFGF. Embryos were injected with mRNA for *globin*, *ndk*, or XFD into the animal pole region at the 2-cell stage. Animal caps were dissected at the blastula stage (stages 8.5 to 9), treated with 100 ng ml⁻¹ bFGF until the sibling embryos reached stage 11, and assayed for β -gal staining and *bra* expression by whole-mount *in situ* hybridization. Only animal caps stained in red by β -gal were scored. Amounts of mRNA (pg per embryo): *globin* (negative control) and *ndk*, 1,000; XFD, 200.

Methods

Organisms

We used a clonal strain of the planarian, *D. japonica*, obtained from the Irima river, Gifu prefecture, Japan, established in Watanabe's laboratory in the Himeji Institute of Technology. Intact animals were cultured in autoclaved tap water at 21 °C, and they were starved for 2 weeks before being used. For all experiments planarians 4–7 mm in length were used.

Whole-mount *in situ* hybridization

Planarians that had been starved for two weeks were used. Animals were treated with 2% HCl for 5 min at 4 °C and then fixed in Carnoy's solution (ethanol:chloroform:acetic acid at 6:3:1, respectively) for 2 h at 4 °C. Hybridization was carried out using 20 ng ml⁻¹ of digoxigenin (DIG)-labelled riboprobes, as previously described^{5,12}.

Synthesis of dsRNA

This was done basically as previously described^{10,16}. pBluescriptII SK + containing a cDNA insert of 2,483 base pairs (bp) including a putative full-length open reading frame (ORF) of *ndk* was digested with *SacI* or *KpnI* to synthesize antisense (T7) or sense (T3) RNAs, respectively. After RNA synthesis, the samples were digested with DNaseI for 20 min at 37 °C; they were then pooled and extracted with phenol/chloroform. The RNAs were denatured for 20 min at 65 °C, and annealed for 40 min at 37 °C. After ethanol precipitation, dsRNA was resuspended in H₂O that had been treated with diethyl pyrocarbonate (DEPC). Electrophoretic mobilities of dsRNA and single-strand RNA were assessed in 1.5% agarose gels. For double and triple-gene injections, dsRNAs were individually synthesized and then mixed, precipitated with ethanol and resuspended in 10 μ l DEPC-treated H₂O.

Analysis of endogenous expression of *ndk* after dsRNA injection

Clone 721HH containing *ndk* was digested with *EcoRI* to generate two non-overlapping fragments of about 1 kb (5' region) and 1.4 kb (3' region). The 1-kb fragment was re-cloned in *EcoRI* site dephosphorylated pBluescript SK + and the 1.4-kb fragment was self-ligated. After subcloning, the 1-kb fragment was digested with *SacI* or *KpnI* to synthesize sense (T7) or antisense (T3) RNA, respectively for dsRNA synthesis. The 1.4-kb fragment was digested with *SacI* to synthesize antisense (T7) for the DIG-RNA probe.

Microinjection and amputation

Intact planarians were injected with dsRNA three times (32 nl per injection) for three consecutive days using a Drummond Scientific (Broomall) Nanoject injector¹⁰. Control animals were injected with H₂O. Two–three hours after the last injections, animals were amputated at different levels along the antero–posterior axis using sterile surgical blades (Keisei Medical Industrial Co.). Three kinds of regenerating fragments were obtained after amputation: head fragments capable of regenerating new central and tail regions; trunk fragments able to regenerate new head and tail regions; and tail fragments capable of regenerating a new head and central regions. During the regeneration process, the samples were maintained at 21 °C.

Immunofluorescence

This was carried out as previously described¹⁰. Intact and regenerating planarians were treated with 2% HCl for 5 min at 4 °C and then fixed in Carnoy's solution for 3 h at 4 °C. The samples were rinsed in 100% methanol, and bleached overnight at room temperature in 6% H₂O₂ in methanol. After bleaching, the samples were rehydrated through a graded methanol series (75%, 50% and 25% in PBS-0.3% Triton X-100) for 15 min each, and rinsed in PBS containing 0.3% triton X-100 (PBST; Sigma). After blocking for 2 h in PBST containing 0.25% BSA (PBSTB; Sigma) samples were incubated in PBSTB at room temperature for 14–16 h with the following primary antibodies: a polyclonal antibody against planarian synaptotagmin (SYT)²⁷, and a monoclonal antibody against visual cells (VC-1)²⁸ diluted 1:2,000 and 1:10,000, respectively. After washing in PBSTB for 6–8 h with several changes of the medium, the samples were incubated in secondary anti-mouse Alexa 488 diluted 1:400 in PBSTB, overnight at room temperature. Finally, the samples were washed for several hours in PBST, mounted in DAKO fluorescent mounting medium and observed with a Olympus BX62 microscope.

X-ray irradiation

Intact planarians were irradiated using a Softex B-4 X-ray source operating at 18 kV, 5 mA as previously described¹⁴. Four days after irradiation, the animals were injected with *ndk* dsRNA as described above. Twelve days after injections, they were fixed for *in situ* hybridization experiments.

mRNA injection experiments with *Xenopus* embryos

Fertilization, manipulation and mRNA injection of *Xenopus* embryos were as previously described²⁹. The entire or truncated coding region of *Dugesia ndk* was amplified by PCR and inserted into the *Clal* and *XbaI* sites of pCS2 + to construct pCS2 + *ndk*, pCS2 + *ndk*-ΔC1 (amino acids 1–408), and pCS2 + *ndk*-ΔC2 (amino acids 1–362). Whole-mount *in situ* hybridization was performed with *Xenopus bra* probe²⁰ as previously described³⁰.

Received 4 April; accepted 15 July 2002; doi:10.1038/nature01042.

- Pearson, H. The regeneration gap. *Nature* **414**, 388–390 (2001).
- Baguña, J., Saló, E. & Auladell, C. Regeneration and pattern formation in planarians. III. Evidence that neoblasts are totipotent stem cells and the source of blastema cells. *Development* **107**, 77–86 (1989).
- Agata, K. & Watanabe, K. Molecular and cellular aspects of planarian regeneration. *Semin. Cell Dev. Biol.* **10**, 77–83 (1999).
- Newmark, P. A. & Sánchez Alvarado, A. Not your father's planarian: a classic model enters the era of functional genomics. *Nature Rev. Genet.* **3**, 210–219 (2002).
- Agata, K. *et al.* Structure of the planarian nervous system (CNS) revealed by neuronal cell markers. *Zool. Sci.* **15**, 433–440 (1998).
- Cebrià, F. *et al.* Dissecting planarian CNS regeneration by the expression of neural-specific genes. *Dev. Growth Differ.* **44**, 135–146 (2002).
- Kato, K., Orii, H., Watanabe, K. & Agata, K. The role of dorsoventral interaction in the onset of planarian regeneration. *Development* **126**, 1031–1040 (1999).
- Kobayashi, C., Nogi, T., Watanabe, K. & Agata, K. Ectopic pharynxes arise by regional reorganization after anterior/posterior chimera in planarians. *Mech. Dev.* **89**, 25–34 (1999).
- Orii, H. *et al.* The planarian HOM/HOX homeobox genes (Plox) expressed along the anteroposterior axis. *Dev. Biol.* **210**, 456–468 (1999).
- Sánchez Alvarado, A. & Newmark, P. A. Double-stranded RNA specifically disrupts gene expression during planarian regeneration. *Proc. Natl Acad. Sci. USA* **96**, 5049–5054 (1999).
- Shibata, N. *et al.* Expression of *vasa* (*vas*)-related genes in germline cells and totipotent somatic stem cells of planarians. *Dev. Biol.* **206**, 73–87 (1999).
- Umesono, Y., Watanabe, K. & Agata, K. Distinct structural domains in the planarian brain defined by the expression of evolutionarily conserved homeobox genes. *Dev. Genes Evol.* **209**, 31–39 (1999).
- Cebrià, F. *et al.* The expression neural-specific genes reveals the structural and molecular complexity of the planarian central nervous system. *Mech. Dev.* **116**, 199–204 (2002).
- Okawa, K. *et al.* Planarian FGFR homologues expressed in stem cells and cephalic ganglions. *Dev. Growth Differ.* **44**, 191–204 (2002).
- Wiedemann, M. & Trueb, B. Characterization of a novel protein (FGFRL1) from human cartilage related to FGF receptors. *Genomics* **69**, 275–279 (2000).
- Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811 (1998).
- Amaya, E., Musci, T. J. & Kirschner, M. W. Expression of a dominant negative mutant of the FGF receptor disrupts mesoderm formation in *Xenopus* embryos. *Cell* **66**, 257–270 (1991).
- Conlon, F. L. & Smith, J. C. Interference with brachyury function inhibits convergent extension, causes apoptosis, and reveals separate requirements in the FGF and activin signalling pathways. *Dev. Biol.* **213**, 85–100 (1999).
- Smith, J. C., Price, B. M., Green, J. B., Weigel, D. & Herrmann, B. G. Expression of a *Xenopus* homolog of Brachyury (T) is an immediate-early response to mesoderm induction. *Cell* **67**, 79–87 (1991).
- Isaacs, H. V., Pownall, M. E. & Slack, J. M. eFGF regulates *Xbra* expression during *Xenopus* gastrulation. *EMBO J.* **13**, 4469–4481 (1994).
- Onichtchouk, D. *et al.* Silencing of TGF- β signalling by the pseudoreceptor BAMBI. *Nature* **401**, 480–485 (1999).
- Tsang, M., Friesel, R., Kudoh, T. & Dawid, I. B. Identification of Sef, a novel modulator of FGF signalling. *Nature Cell Biol.* **4**, 165–169 (2002).
- Fürthaus, M., Lin, W., Ang, S. L., Thisse, B. & Thisse, C. Sef is a feedback-induced antagonist of Ras/MAPK-mediated FGF signalling. *Nature Cell Biol.* **4**, 170–174 (2002).
- Dowd, C. J., Cooney, C. L. & Nugent, M. A. Heparan sulfate mediates bFGF transport through basement membrane by diffusion with rapid reversible binding. *J. Biol. Chem.* **274**, 5236–5244 (1999).
- Launay, C., Fromentoux, V., Shi, D. L. & Boucaut, J. C. A truncated FGF receptor blocks neural induction by endogenous *Xenopus* inducers. *Development* **122**, 869–880 (1996).
- Streit, A., Berliner, A. J., Papanayotou, C., Sirulnik, A. & Stern, C. D. Initiation of neural induction by FGF signalling before gastrulation. *Nature* **406**, 74–78 (2000).
- Tazaki, A. *et al.* Neural network in planarian revealed by an antibody against planarian synaptotagmin homologue. *Biochem. Biophys. Res. Commun.* **260**, 426–432 (1999).
- Sakai, F., Agata, K., Orii, H. & Watanabe, K. Organization and regeneration ability of spontaneous supernumerary eyes in planarians—Eye regeneration field and pathway selection by optic nerves. *Zool. Sci.* **17**, 375–381 (2000).
- Taira, M., Otani, H., Saint-Jeanet, J. P. & Dawid, I. B. Role of the LIM class homeodomain protein *Xlim-1* in neural and muscle induction by the Spemann organizer in *Xenopus*. *Nature* **372**, 677–679 (1994).
- Harland, R. M. in *Methods in Cell Biology* (eds Kay, B. K. & Peng, H. B.) **36**, 685–695 (Academic, San Diego, 1991).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements We thank S. Kuratani for comments on the manuscript, J. Brookes and T. Miyata for promoting collaboration, and D. Turner for pCS2 + and pCS2 + n3-gal. This work was supported by Special Coordination Funds for Promoting Science and Technology (K.A.), a Grant-in-Aid for Creative Basic Research (K.A. and T.G.), a Grant-in-Aid for Scientific Research on Priority Areas (K.A. and M.T.) and the National Institutes of Health National Institute of General Medical Sciences (A.S.A.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to K.A. (e-mail: agata@cdb.riken.go.jp). The *Dugesia ndk* cDNA sequence has been deposited in DDBJ under accession number AB071948.

The polycomb group protein EZH2 is involved in progression of prostate cancer

Sooryanarayana Varambally^{*†}, Saravana M. Dhanasekaran^{*†}, Ming Zhou^{*}, Terrence R. Barrette^{*}, Chandan Kumar-Sinha^{*}, Martin G. Sanda^{‡§}, Debashis Ghosh^{||}, Kenneth J. Pienta^{‡§¶}, Richard G. A. B. Sewalt[#], Arie P. Otte[#], Mark A. Rubin^{*‡§} & Arul M. Chinnaiyan^{*‡§}

^{*} Departments of Pathology, [‡] Urology, ^{||} Biostatistics and [¶] Internal Medicine, and [§] Comprehensive Cancer Center, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA

[#] Swammerdam Institute for Life Sciences, BioCentrum Amsterdam, University of Amsterdam, 1018 TV Amsterdam, The Netherlands

[†] These authors contributed equally to this work

Prostate cancer is a leading cause of cancer-related death in males and is second only to lung cancer. Although effective surgical and radiation treatments exist for clinically localized prostate cancer, metastatic prostate cancer remains essentially incurable. Here we show, through gene expression profiling¹, that the polycomb group protein enhancer of zeste homolog 2 (EZH2)^{2,3} is over-

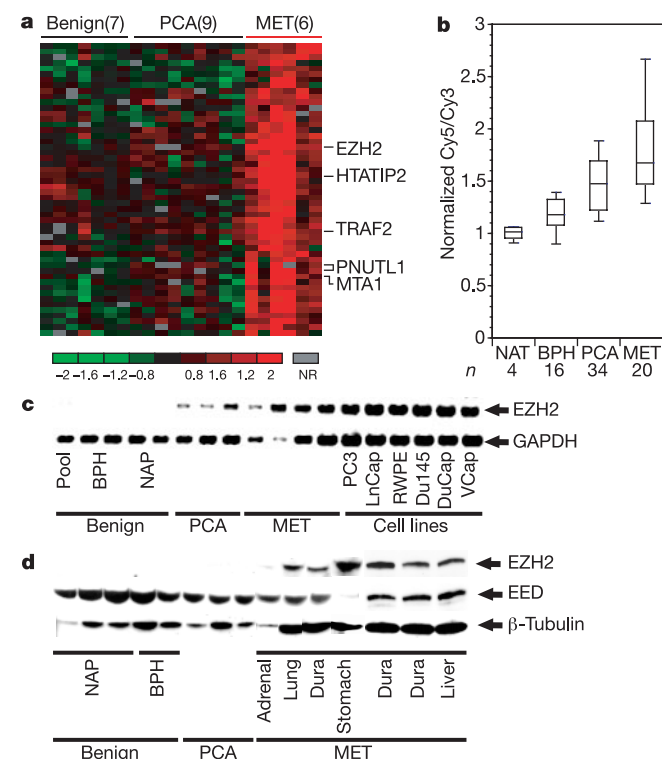


Figure 1 Overexpression of EZH2 in metastatic hormone-refractory prostate cancer (MET). **a**, Cluster diagram depicting genes that distinguish MET from clinically localized prostate cancer (PCA). Genes upregulated in METs relative to prostate cancer are shown. Red and green represent upregulation and downregulation, respectively, relative to the median of the reference pool. Grey represents technically inadequate or missing data, and black represents equal expression relative to the reference sample. **b**, DNA microarray analysis of prostate cancer shows upregulation of EZH2 in METs. BPH, benign prostatic hyperplasia; NAT, normal adjacent prostate tissue. **c**, RT-PCR analysis of the EZH2 transcript in prostate tissue and cell lines. **d**, Increased expression of EZH2 protein in METs. Immunoblot analysis of EZH2 and EED in prostate tissue extracts. The site of metastasis is indicated.

expressed in hormone-refractory, metastatic prostate cancer. Small interfering RNA (siRNA) duplexes⁴ targeted against EZH2 reduce the amounts of EZH2 protein present in prostate cells and also inhibit cell proliferation *in vitro*. Ectopic expression of EZH2 in prostate cells induces transcriptional repression of a specific cohort of genes. Gene silencing mediated by EZH2 requires the SET domain and is attenuated by inhibiting histone deacetylase activity. Amounts of both EZH2 messenger RNA and EZH2 protein are increased in metastatic prostate cancer; in addition, clinically localized prostate cancers that express higher concentrations of EZH2 show a poorer prognosis. Thus, dysregulated expression of EZH2 may be involved in the progression of prostate cancer, as well as being a marker that distinguishes indolent prostate cancer from those at risk of lethal progression.

Perturbation of the transcriptional 'memory' of a cell can lead to severe developmental defects^{5,6}. Lack of differentiation, or anaplasia, is a hallmark of cancer that results from normal cells 'forgetting' their cellular identity. Thus, it is not surprising that dysregulation of the transcriptional maintenance system can lead to malignancy⁶⁻⁸. Two groups of proteins have been implicated in maintaining homeotic gene expression and include polycomb (PcG) and trithorax (trxG) group proteins^{9,10}. PcG proteins act in large complexes and are thought to repress gene expression, whereas trxG proteins are operationally defined as antagonists of PcG proteins and thus activate gene expression^{6,9}. In human malignancies, PcG and trxG proteins have been found to be dysregulated mainly in cells of haematopoietic origin¹¹⁻¹³. EZH2 is the human homologue of the

Drosophila protein Enhancer of Zeste (E(z))², which genetic data define as a PcG protein with additional trxG properties¹⁰. E(z) and EZH2 contain a SET domain, a highly conserved domain found in chromatin-associated regulators of gene expression that often modulate cell growth pathways¹⁴.

We previously reported gene expression profiles of benign prostate, clinically localized prostate cancer and metastatic prostate cancer¹. To characterize genes that specifically mark the molecular transition from organ-confined prostate cancer to metastatic prostate cancer, we implemented a statistical technique called significance analysis of microarrays (SAM)¹⁵. This approach identified 55 genes that were significantly upregulated (Fig. 1a) and 480 genes that were significantly downregulated (Supplementary Information) in metastatic prostate cancer relative to localized prostate cancer (at a false discovery rate (FDR) of 0.9% on microarrays containing 10,000 genes). Notably, the gene at the top of the 'list' of upregulated genes in metastatic prostate cancer was EZH2, with a *d*-score¹⁵ of 4.58 and a gene-specific FDR of 0.0012 (see Supplementary Information for the complete SAM analysis). Fig. 1b summarizes the gene expression of EZH2 in 74 prostate tissue specimens analysed on DNA microarrays¹. The EZH2 transcript was significantly increased in metastatic prostate cancer with respect to clinically localized prostate cancer (Mann-Whitney test, $P = 0.001$) and benign prostate ($P = 0.0001$).

To validate the DNA microarray results, we carried out polymerase chain reaction with reverse transcription (RT-PCR) on 18 prostate samples and cell lines. As expected, the levels of EZH2 mRNA were increased in malignant relative to benign prostate

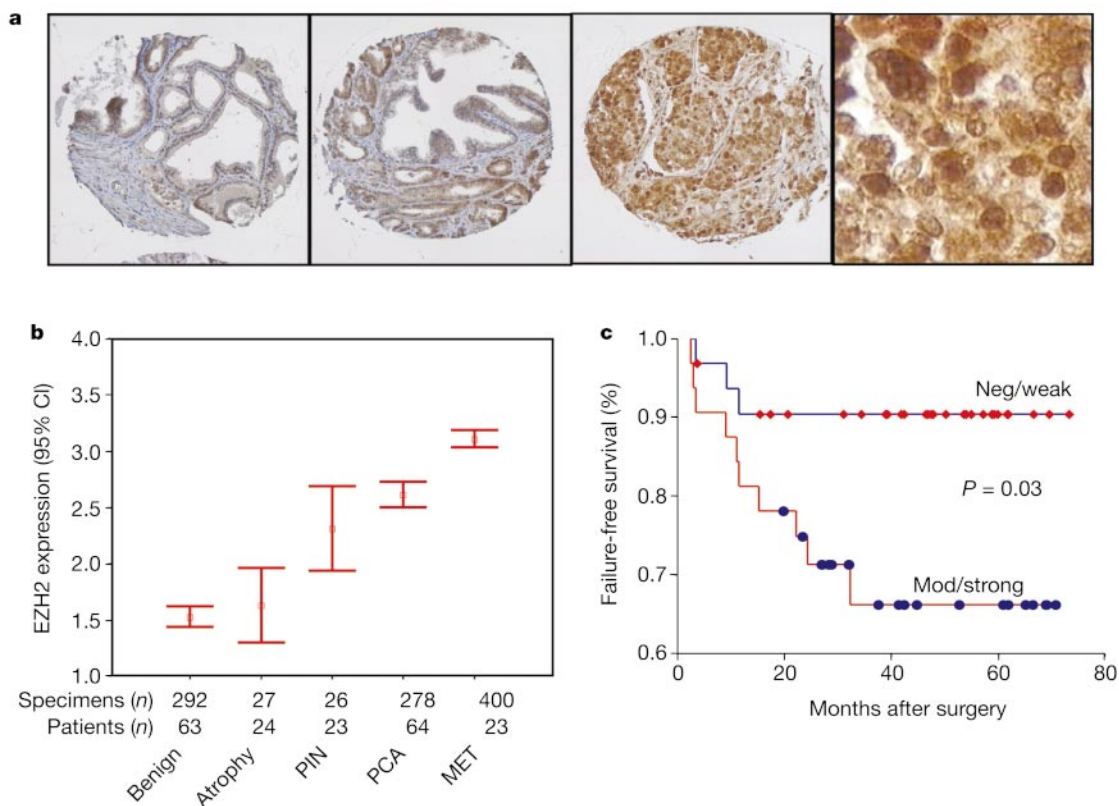


Figure 2 Amounts of EZH2 protein correlate with the aggressiveness of prostate cancer. **a**, Representative tissue microarray elements stained with antibody to EZH2. Immunohistochemical stains show weak or absent nuclear staining of benign prostate cancer (left, original magnification $\times 100$), and strong nuclear staining in metastatic hormone-refractory prostate cancers (centre right and right, original magnification $\times 100$ and $\times 1,000$, respectively). Also shown is a benign prostate adjacent to prostate

cancer (centre left, original magnification $\times 100$). **b**, Tissue microarray analysis of EZH2 expression. The mean EZH2 protein expression for the indicated prostate tissues is summarized using error bars with 95% confidence intervals. **c**, Kaplan-Meier analysis shows that individuals with clinically localized prostate cancer that have high expression of EZH2 (moderate to strong staining) have a greater risk for recurrence of prostate cancer after prostatectomy (log rank test, $P = 0.03$).

samples (Fig. 1c). We also analysed tissue extracts by immunoblotting. The amounts of EZH2 protein were markedly increased in metastatic prostate cancer relative to localized prostate cancer or benign prostate (Fig. 1d). Notably, EED, a PcG protein that forms a complex with EZH2 (refs 16, 17), did not show similar protein dysregulation.

We evaluated the expression of EZH2 protein in a wide range of prostate tissues ($n = 1,023$ tissue microarray elements) to determine the extent of its expression *in situ* (Fig. 2a, b). Of these samples, 400 were from 23 individuals who died of hormone-refractory metastatic prostate cancer¹⁸. When highly expressed, EZH2 was observed mainly in the nucleus (Fig. 2a, right), as suggested previously¹¹. The intensity of EZH2 staining increased from benign, prostatic atrophy, prostatic intraepithelial neoplasia, to clinically localized prostate cancer, with a respective median staining intensity of 1.5 (standard error (s.e.) 0.04, 95% confidence interval (CI) 1.4–1.6), 1.6 (s.e. 0.2, 95%CI 1.3–2.0), 2.3 (s.e. 0.2, 95%CI 1.9–2.7) and 2.6 (s.e. 0.1, 95%CI 2.5–2.7), respectively (Fig. 2b). The strongest EZH2 expression was in metastatic prostate cancer, with a median staining intensity of 3.1 (s.e. 0.03, 95%CI 3.0–3.2). There was a statistically significant difference in EZH2 staining intensity between benign prostate tissue and localized prostate cancer (analysis of variance (ANOVA) post-hoc analysis mean difference 1.0, $P < 0.0001$). Likewise, metastatic prostate cancer had significantly higher expression of EZH2 than did clinically localized prostate cancer (ANOVA post-hoc analysis mean difference 0.5, $P < 0.0001$). These findings suggested that as prostate neoplasia progresses there is a trend towards increased expression of EZH2 protein. They also suggested that EZH2 concentrations might indicate how aggressive an individual's prostate cancer is, given that the highest expression was observed in hormone-refractory, metastatic prostate cancer.

We therefore examined whether EZH2 protein could be used to predict clinical outcome in men treated with surgery for clinically

localized prostate cancer. Using a previously validated tissue microarray¹⁹, we examined 278 specimens from 64 individuals with clinical follow-up. To test the utility of EZH2 as a potential tissue biomarker for prostate cancer, we examined the clinical outcome of these 64 cases, taking into account clinical and pathological parameters. Clinical failure was defined as either an increase of 0.2 ng ml^{-1} prostate-specific antigen (PSA) or recurrence of disease after prostatectomy (for example, development of metastatic disease).

By Kaplan–Meier analysis (Fig. 2c), an EZH2 staining intensity of 3 or higher was associated significantly with clinical failure in 31% (10/32) of individuals (log rank $P = 0.03$), whereas a staining intensity of less than 3 was found in 9% (3/32) of individuals with clinical failure. There was no significant correlation between EZH2 amounts and Gleason score, tumour stage or surgical margin status. Notably, there was a significant ($P = 0.048$) albeit weak (Pearson coefficient = 0.33) correlation between EZH2 protein and proliferation index *in situ*, as assessed by the Ki-67 labelling index (Supplementary Information). Multivariable Cox hazards regression analysis showed that EZH2 protein expression (≥ 3 versus < 3) was the best predictor of clinical outcome with a recurrence ratio of 4.6 (95%CI 1.2–17.1, $P = 0.02$), which was significantly better than surgical margin status, maximum tumour dimension, Gleason score and pre-operative PSA (Supplementary Information). Thus, monitoring the amounts of EZH2 protein in prostate specimens may provide additional prognostic information that is not discernible with current clinical and pathology parameters alone.

To examine the role of EZH2 in prostate cancer progression, we used RNA interference (RNAi) to disrupt EZH2 expression in transformed prostate cells *in vitro*. We designed duplexes of 21-nucleotide RNA (siRNAs) to target EZH2 (ref. 4) and tested them on the transformed androgen-responsive prostate cell line RWPE^{20,21} and the metastatic prostate cancer cell line PC3. After

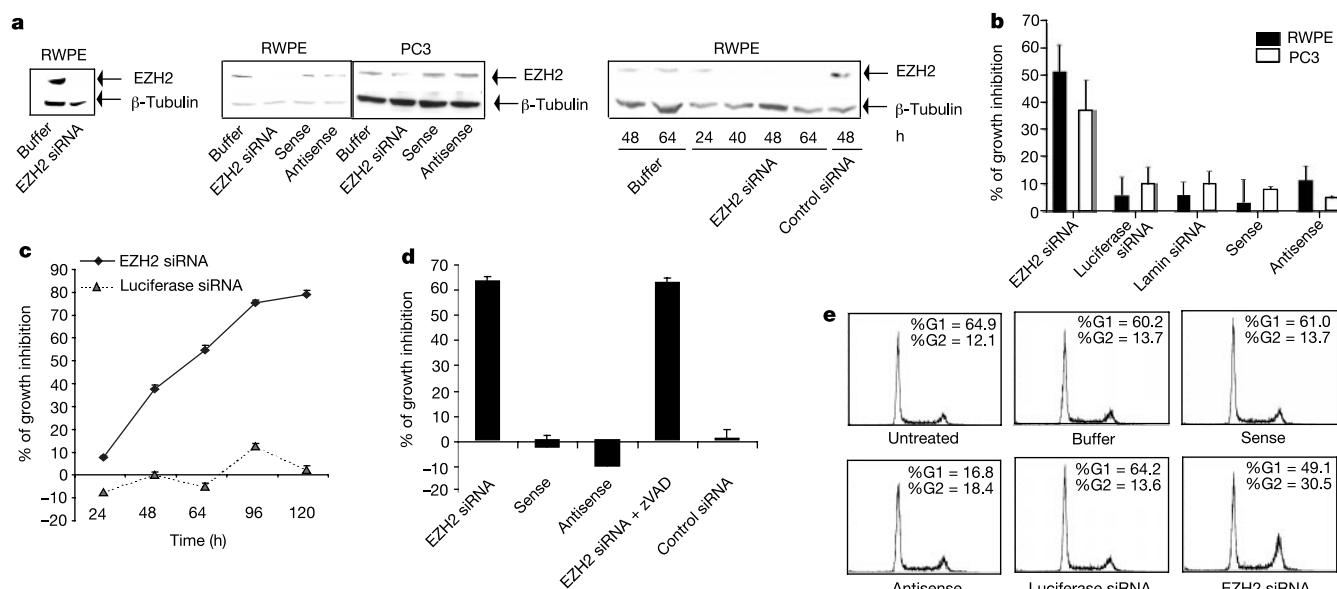


Figure 3 A role for EZH2 in prostate cell proliferation. **a**, Immunoblot analysis of prostate cells (RWPE, PC3) incubated with EZH2 siRNA duplexes. Sense and antisense oligonucleotides of EZH2 and unrelated siRNA duplexes were used as controls. **b**, RNAi of EZH2 decreases cell proliferation, as assessed by cell counting assay after 48 h of incubation with EZH2 siRNA. Significant inhibition of growth is seen in cells treated with EZH2 siRNA relative to those treated with unrelated controls (Student's *t*-test, $P < 0.001$). Data are from two independent experiments carried out in quadruplicate.

c, Time course of growth inhibition by EZH2 siRNA. RWPE cells were treated with EZH2 or luciferase siRNA for indicated time periods and the cells were counted. **d**, RNAi of EZH2 inhibits cell proliferation, as assessed by WST assay. EZH2 siRNA significantly inhibited cell proliferation as compared with controls (Student's *t*-test, $P < 0.001$). **e**, RNAi of EZH2 induces cell-cycle arrest of prostate cells as assessed by flow cytometry.

48 h of transfection with siRNA duplexes, quantification of the amounts of endogenous EZH2 protein showed a specific down-regulation of protein in prostate cell lines (Fig. 3a). Sense or antisense oligonucleotides, as well as unrelated siRNA duplexes, did not affect the amounts of EZH2 (Fig. 3a, middle and right), verifying the specificity of the siRNA approach.

We next examined the phenotype of the prostate cells with reduced EZH2 expression. RWPE cell counts taken 48 h after transfection with siRNA showed a marked (62%) inhibition of cell growth caused by the EZH2 siRNA duplex, whereas the corresponding sense and antisense EZH2 oligonucleotides or unrelated duplexes showed minimal inhibition (Fig. 3b). The PC3 prostate cancer cell line showed a similar growth inhibition by EZH2 siRNA, indicating that these findings were not specific to the RWPE cell line (Fig. 3b). The effect of EZH2 siRNA on prostate cell growth was monitored at time points ranging from 24 to 120 h (Fig. 3c), during which over 80% growth inhibition was observed.

Using a commercially available cell proliferation reagent WST-1, which measures mitochondrial dehydrogenase activity, we observed a decrease in cell proliferation in cells transfected with the EZH2 siRNA duplex but not with the unrelated duplexes (Fig. 3d). In the time frame considered (48 h), RNAi of EZH2 did not induce apoptosis, as assessed by propidium iodide staining of nuclei or cleavage of poly(adenosine diphosphate)-ribosylpolymerase (data not shown). Consistent with this, the broad-spectrum caspase inhibitor, zVAD-fmk, failed to attenuate the inhibition of cell proliferation induced by EZH2 siRNA (Fig. 3d). Notably, flow cytometric analysis of prostate cells treated with EZH2 siRNA showed cell-cycle arrest in the G2/M phase (Fig. 3e). Unrelated control siRNA duplexes failed to induce a similar dysregulation of the cell cycle. Together, these observations indicate that EZH2 may be involved in the proliferation of prostate cells by mitigating the G2/M transition.

To understand further the functional role of EZH2 in prostate cells, we generated an epitope-tagged version of wild-type EZH2 and a deletion mutant of EZH2 lacking the conserved SET domain¹⁴ in the eukaryotic expression vector, pcDNA3 (Fig. 4a, b). We also designed an 'inducible' version^{22,23} of EZH2 by creating a fusion protein with a modified murine oestrogen receptor (ER; Fig. 4a, b).

PcG proteins have been proposed to mediate their functions by repressing target genes^{2,5}. To test this hypothesis, we transiently transfected RWPE prostate cells with wild-type EZH2 and monitored alterations in global gene expression using DNA microarrays. The RNA from the transfected cell line was labelled with one fluorescent dye, whereas the paired reference sample was labelled with a second distinguishable fluorescent dye. By making direct comparisons between 'gene'-transfected cell lines and cells lines transfected with control vector, we emphasized the molecular differences between the samples and obviated the need for complete transfection efficiency.

When EZH2 was overexpressed in RWPE cells or SUM149 breast carcinoma cells, there was a consistent repression of a cohort of genes (Fig. 4c, d). When compared with vector-transfected cells, the only gene that was upregulated significantly in EZH2-transfected cells was EZH2 itself (Fig. 4c). The EZH2-mediated transcriptional repression was dependent on an intact SET domain (Fig. 4c), because deletion of this domain did not produce a repressive phenotype and in some cases 'de-repressed' genes. EZH2 interacts with histone deacetylase 2 (HDAC2) through the EED protein²⁴, and in our system EZH2-mediated gene silencing was dependent on HDAC activity, because the commonly used HDAC inhibitor trichostatin A (TSA) completely abrogated the effects of EZH2 (Fig. 4c). Thus, EZH2 function requires both an intact SET domain and endogenous HDAC activity.

To identify genes that are significantly repressed by EZH2, we compared cells transfected with wild-type EZH2 with cells trans-

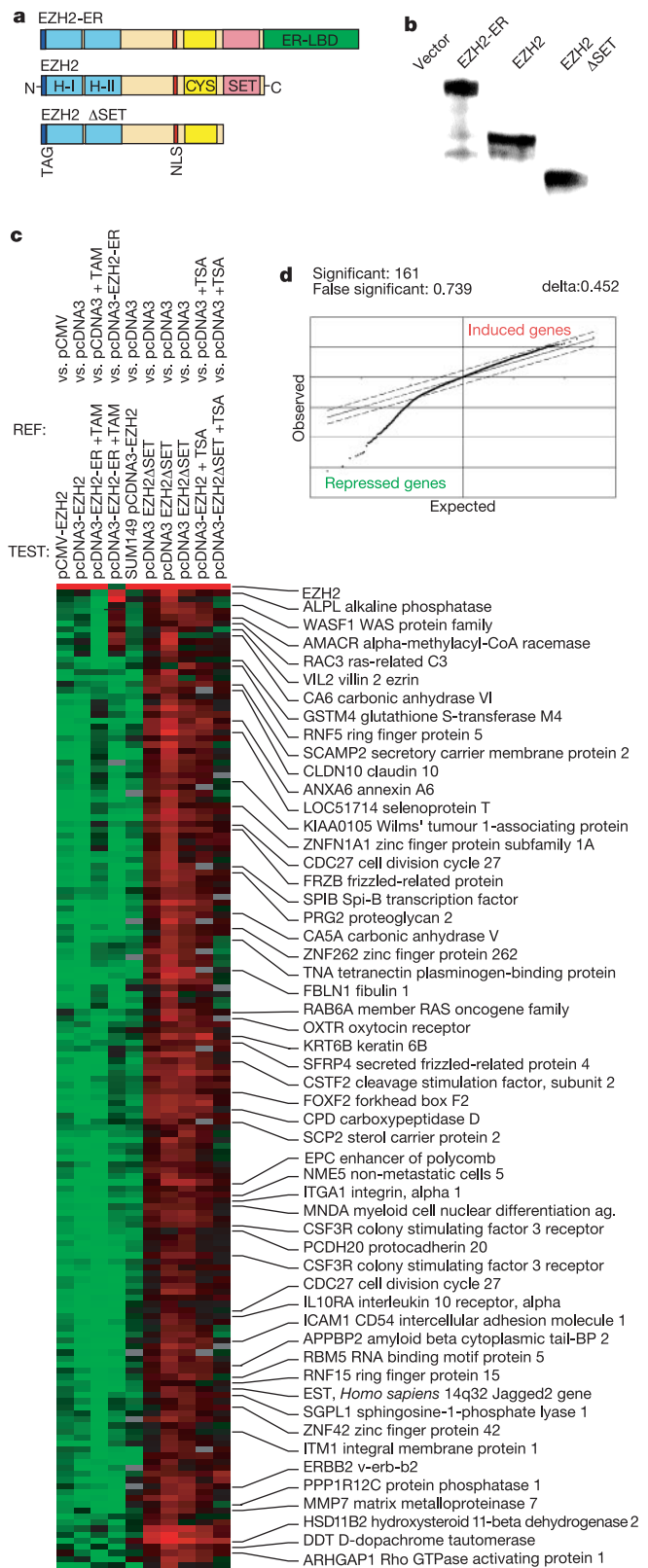


Figure 4 EZH2 functions as a transcriptional repressor in prostate cells. **a**, EZH2 constructs. CYS, cysteine-rich domain; ER, ligand-binding domain of the oestrogen receptor; H-I and H-II, homology domains that share similarity with E(z); NLS, nuclear localization signal; SET, SET domain; TAG, Myc epitope tag. **b**, Expression of the EZH2 constructs. **c**, Cluster diagram of genes significantly repressed by overexpression of EZH2. See Fig. 1a for the matrix colour scheme and Supplementary Information for the complete data set. TAM, tamoxifen; TSA, trichostatin A. **d**, SAM analysis¹⁵ of gene expression profiles of cells transfected with EZH2 versus cells transfected with EZH2ΔSET.

fected with EZH2ΔSET. We found that 163 genes were consistently repressed, whereas no genes were activated at an FDR of 0.0045 (Fig. 4d, and Supplementary Information). The list of genes that were repressed indicates several intriguing associations, including EZH2-mediated repression of specific PcG proteins, transcription factors and cell-cycle regulators (Supplementary Information).

In summary, we have identified the association of EZH2 with advanced prostate cancer by gene expression profiling of tumour specimens from individuals who died of metastatic disease. Measuring EZH2 amounts (in prostate cancer specimens) might have potential as a molecular determinant of prostate cancer progression and metastasis. In addition, we have shown that EZH2 has a role in mediating cell proliferation and transcriptional repression in prostate cells. Overall, our study indicates that dysregulation of the transcriptional memory machinery may contribute to the lethal progression of prostate cancer and may provide a potential mechanism for the constellation of genes repressed in metastatic disease. □

Methods

SAM analysis of prostate cancer gene expression

We carried out SAM analysis by comparing gene expression profiles of metastatic prostate cancer samples with those of clinically localized prostate cancer samples from our previous work¹. Genes were analysed using Cluster²⁵, implementing average linkage hierarchical clustering of genes, and the output (Supplementary Information) was visualized by Treeview²⁵.

RT-PCR

We carried out RT-PCR amplifications with 1 µg of total RNA isolated from the indicated prostate tissues and cell lines. Primer sequences are given in the Supplementary Information.

Immunoblot analysis

Prostate tissue extracts were separated by SDS-PAGE and blotted onto nitrocellulose membranes. Antibodies against EZH2 and EED (ref. 17) and a polyclonal antibody against β-tubulin (Santa Cruz) were used at 1:1,000 dilution for immunoblot analysis.

Tissue microarray analysis

The clinically stratified prostate cancer tissue microarrays used in this study have been described^{1,19,26}. Tissues were from the radical prostatectomy series at the University of Michigan and from the Rapid Autopsy Program, which are both part of Michigan Prostate SPORE Tissue Core. We obtained Institutional Review Board approval to procure and analyse the tissues used in this study.

We evaluated EZH2 protein expression on a wide range of prostate tissue to determine the intensity and extent of expression *in situ*. Immunohistochemistry was carried out on three high-density tissue microarrays containing samples of benign prostate, prostatic atrophy, prostatic intraepithelial neoplasia, clinically localized prostate cancer and metastatic prostate cancer. We used standard biotin-avidin complex immunohistochemistry to evaluate EZH2 protein expression using an antibody against EZH2. Protein expression was scored as negative (score = 1), weak (2), moderate (3) and strong (4). Four replicate tissue cores were sampled from each of the selected tissue types. EZH2 expression was evaluated in a blind manner separately by M.A.R. and M.Z. using a validated web-based tool^{26,27}, and the median value of all measurements from a single individual was used for subsequent analysis. Staining assessment was highly reproducible between the two pathologists with a κ value of 0.73.

Clinical outcomes analysis

To assess individual variables for risk of recurrence, we used Kaplan-Meier survival analysis and created a univariate Cox proportional hazards model. PSA recurrence was defined as greater than 0.2 ng ml⁻¹ after radical prostatectomy. Covariates included Gleason sum, preoperative PSA, maximum tumour dimension, tumour stage and surgical margin status. To assess the influence of several variables simultaneously, including EZH2 protein expression, we developed a final multivariate Cox proportional hazards model of statistically significant covariates. Statistical significance in univariate and multivariate Cox models was determined by Wald's test. A P value of less than 0.05 was considered statistically significant.

EZH2 constructs

Myc-tagged EZH2-pCMV was a gift from T. Jenuwein. The Myc-EZH2 fragment was subcloned into the expression vector pCDNA3 (Invitrogen). The EZH2-ER and EZH2ΔSET constructs are described in the Supplementary Information.

RNAi

We chemically synthesized 21-nucleotide sense and antisense RNA oligonucleotides (Dharmacon) and annealed them to form duplexes. The EZH2 siRNA was targeted to the region corresponding to residues 85–106 of human EZH2 (NM004456). Control siRNA duplexes targeted luciferase, lamin and AMACR (NM014324). The human transformed prostate cell lines RWPE²⁰ and PC3 were plated at 2 × 10⁵ cells per well in a 12-well plate

for immunoblot analysis, cell counts and FACS analysis, and at 1.5 × 10⁴ cell per well in a 96-well plate for WST-1 proliferation assays. Twelve hours after plating, the cells were transfected with 60 pmol of siRNA duplex, sense or antisense oligonucleotides using oligofectamine (Invitrogen). We carried out a second identical transfection 24 h later. At certain time points after the first transfection, the cells were lysed for immunoblot analysis and treated with trypsin for estimating cell numbers or FACS analysis. For cell counts at 96 and 120 h, the cells were treated with trypsin and replated in 6-well dishes 64 h after the first transfection.

Cell proliferation assays

Cell proliferation was determined by a colorimetric assay of cell viability that is based on the cleavage of the tetrazolium salt WST-1 by mitochondrial dehydrogenases (Roche). The absorbance of the formazan dye formed, which correlates with the number of metabolically active cells in the culture, was measured at 450 nm 1 h after adding the reagent. Cell counts were estimated by treating the cells with trypsin, followed by analysis on a Coulter cell counter at specified time points.

Flow cytometric analysis

Trypsin-treated cells were washed with PBS and fixed in 70% ethanol overnight for FACS analysis. Before staining with propidium iodide, the cells were washed again with PBS and analysed by flow cytometry.

Microarray analysis of EZH2-transfected cells

Initial testing of the transient transfection analysis system (data not shown) showed that overexpression of TNFR1 (p55) induced similar expression profiles to those observed after incubating cells with TNF²⁸. Samples expressing the EZH2ΔSET mutant were compared with those expressing EZH2 using the SAM analysis package (ref. 15 and Supplementary Information).

Received 13 May; accepted 31 July 2002; doi:10.1038/nature01075.

1. Dhanasekaran, S. M. *et al.* Delineation of prognostic biomarkers in prostate cancer. *Nature* **412**, 822–826 (2001).
2. Laible, G. *et al.* Mammalian homologues of the Polycomb-group gene Enhancer of zeste mediate gene silencing in *Drosophila* heterochromatin and at *S. cerevisiae* telomeres. *EMBO J.* **16**, 3219–3232 (1997).
3. Satijn, D. P. & Otte, A. P. Polycomb group protein complexes: do different complexes regulate distinct target genes? *Biochim. Biophys. Acta* **1447**, 1–16 (1999).
4. Elbashir, S. M. *et al.* Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**, 494–498 (2001).
5. Jacobs, J. J. & van Lohuizen, M. Cellular memory of transcriptional states by Polycomb-group proteins. *Semin. Cell. Dev. Biol.* **10**, 227–235 (1999).
6. Francis, N. J. & Kingston, R. E. Mechanisms of transcriptional memory. *Nature Rev. Mol. Cell Biol.* **2**, 409–421 (2001).
7. Jacobs, J. J., Kieboom, K., Marino, S., DePinho, R. A. & van Lohuizen, M. The oncogene and Polycomb-group gene *bmi-1* regulates cell proliferation and senescence through the *ink4a* locus. *Nature* **397**, 164–168 (1999).
8. Jacobs, J. J. *et al.* Bmi-1 collaborates with c-Myc in tumorigenesis by inhibiting c-Myc-induced apoptosis via INK4a/ARF. *Genes Dev.* **13**, 2678–2690 (1999).
9. Mahmoudi, T. & Verrijzer, C. P. Chromatin silencing and activation by Polycomb and trithorax group proteins. *Oncogene* **20**, 3055–3066 (2001).
10. LaJeunesse, D. & Shearn, A. E(z): a polycomb group gene or a trithorax group gene? *Development* **122**, 2189–2197 (1996).
11. Raaphorst, F. M. *et al.* Coexpression of BMI-1 and EZH2 polycomb group genes in Reed-Sternberg cells of Hodgkin's disease. *Am. J. Pathol.* **157**, 709–715 (2000).
12. van Lohuizen, M. *et al.* Identification of cooperating oncogenes in Eμ-myc transgenic mice by provirus tagging. *Cell* **65**, 737–752 (1991).
13. Haupt, Y., Alexander, W. S., Barri, G., Klinken, S. P. & Adams, J. M. Novel zinc finger gene implicated as myc collaborator by retrovirally accelerated lymphomagenesis in Eμ-myc transgenic mice. *Cell* **65**, 753–763 (1991).
14. Jenuwein, T., Laible, G., Dorn, R. & Reuter, G. SET domain proteins modulate chromatin domains in eu- and heterochromatin. *Cell. Mol. Life Sci.* **54**, 80–93 (1998).
15. Tusher, V. G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl Acad. Sci. USA* **98**, 5116–5121 (2001).
16. van Lohuizen, M. *et al.* Interaction of mouse polycomb-group (Pc-G) proteins Enx1 and Enx2 with Eed: indication for separate Pc-G complexes. *Mol. Cell. Biol.* **18**, 3572–3579 (1998).
17. Sewalt, R. G. *et al.* Characterization of interactions between the mammalian polycomb-group proteins Enx1/EZH2 and EED suggests the existence of different mammalian polycomb-group protein complexes. *Mol. Cell. Biol.* **18**, 3586–3595 (1998).
18. Rubin, M. A. *et al.* Rapid ('warm') autopsy study for procurement of metastatic prostate cancer. *Clin. Cancer Res.* **6**, 1038–1045 (2000).
19. Perrone, E. E. *et al.* Tissue microarray assessment of prostate cancer tumour proliferation in African-American and white men. *J. Natl Cancer Inst.* **92**, 937–939 (2000).
20. Webber, M. M., Bello, D., Kleinman, H. K. & Hoffman, M. P. Acinar differentiation by non-malignant immortalized human prostatic epithelial cells and its loss by malignant cells. *Carcinogenesis* **18**, 1225–1231 (1997).
21. Bello, D., Webber, M. M., Kleinman, H. K., Wartinger, D. D. & Rhim, J. S. Androgen responsive adult human prostatic epithelial cell lines immortalized by human papillomavirus 18. *Carcinogenesis* **18**, 1215–1223 (1997).
22. Littlewood, T. D., Hancock, D. C., Danielian, P. S., Parker, M. G. & Evan, G. I. A modified oestrogen receptor ligand-binding domain as an improved switch for the regulation of heterologous proteins. *Nucleic Acids Res.* **23**, 1686–1690 (1995).
23. Juin, P., Hueber, A. O., Littlewood, T. & Evan, G. c-Myc induced sensitization to apoptosis is mediated through cytochrome c release. *Genes Dev.* **13**, 1367–1381 (1999).

24. van der Vlag, J. & Otte, A. P. Transcriptional repression mediated by the human polycomb-group protein EED involves histone deacetylation. *Nature Genet.* **23**, 474–478 (1999).
25. Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA* **95**, 14863–14868 (1998).
26. Manley, S., Mucci, N. R., De Marzo, A. M. & Rubin, M. A. Relational database structure to manage high-density tissue microarray data and images for pathology studies focusing on clinical outcome: the prostate specialized program of research excellence model. *Am. J. Pathol.* **159**, 837–843 (2001).
27. Bova, G. S. *et al.* Web-based tissue microarray image data analysis: initial validation testing through prostate cancer Gleason grading. *Hum. Pathol.* **32**, 417–427 (2001).
28. Kumar-Sinha, C., Varambally, S., Sreekumar, A. & Chinnaiyan, A. M. Molecular cross-talk between the TRAIL and interferon signaling pathways. *J. Biol. Chem.* **277**, 575–585 (2001).

Supplementary Information accompanies the paper on *Nature's* website
(<http://www.nature.com/nature>).

Acknowledgements This study was made possible by tissues donated by patients with metastatic prostate cancer enrolled in the University of Michigan, Rapid Autopsy Program funded by the Specialized Program of Research Excellence (SPORE) in Prostate Cancer at the National Cancer Institute. We thank J. Wei for clinical data collection; K. Hamer for preparing polyclonal antibodies; R. Kunkel for figure preparation; J. Harwood and M. LeBlanc for technical assistance; A. Menon for sequence verification; and C. Ingold and G. Tueckmantel for database assistance. A.M.C. is a Pew Foundation Scholar. This work is supported in part by grants from the NIH (A.M.C.), CaPCURE (A.M.C.) and the Michigan SPORE in Prostate Cancer (K.P., M.A.R., A.M.C. and M.G.S.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.M.C.
(e-mail: arul@umich.edu).

Class IV semaphorin Sema4A enhances T-cell activation and interacts with Tim-2

Atsushi Kumanogoh^{*†}, Satoko Marukawa^{*‡}, Kazuhiro Suzuki^{*†}, Noriko Takegahara^{*}, Chie Watanabe^{*}, EweSeng Ch'ng^{*}, Isao Ishida^{*}, Harutoshi Fujimura[§], Saburo Sakoda[§], Kanji Yoshida^{*} & Hitoshi Kikutani^{*}

^{*} Department of Molecular Immunology, Research Institute for Microbial Diseases, Osaka University, 3-1 Yamada-oka, Suita, Osaka 565-0871, Japan

[‡] Department of Internal Medicine and Molecular Science, and

[§] Department of Neurology, Graduate School of Medicine, Osaka University, 2-2 Yamada-oka, Suita, Osaka 565-0871, Japan

[†] These authors contributed equally to this work

Semaphorins are a family of phylogenetically conserved soluble and transmembrane proteins^{1,2}. Although many soluble semaphorins deliver guidance cues to migrating axons during neuronal development^{3–5}, some members are involved in immune responses^{6–9}. For example, CD100 (also known as Sema4D), a class IV transmembrane semaphorin, signals through CD72 to effect nonredundant roles in immune responses^{7,10–13} in a ligand–receptor system that is distinct from any seen previously in the nervous system^{14,15}. Here we report that the class IV semaphorin Sema4A, which is expressed in dendritic cells and B cells, enhances the *in vitro* activation and differentiation of T cells and the *in vivo* generation of antigen-specific T cells. Treating mice with monoclonal antibodies against Sema4A blocks the development of an experimental autoimmune encephalomyelitis that is induced by an antigenic peptide derived from myelin oligodendrocyte glycoprotein. In addition, expression cloning shows that the Sema4A receptor is Tim-2, a member of the family

of T-cell immunoglobulin domain and mucin domain (Tim) proteins that is expressed on activated T cells.

To study the semaphorins expressed in dendritic cells (DCs), we isolated a complementary DNA fragment of the class IV semaphorin, Sema4A, through cloning based on polymerase chain reaction (PCR) using degenerate oligonucleotide primers derived from motifs that are conserved among members of the semaphorin family. Sema4A was originally identified as semB, the expression of which gradually increases during embryonic development¹⁶, although its functions have not yet been reported. Analysis by PCR with reverse transcription (RT–PCR) showed that there was prominent expression of Sema4A messenger RNA in the brain, spleen, lung, kidney and testis of adult tissues (data not shown).

To investigate the biological function and expression of Sema4A in the immune system, we prepared recombinant soluble mouse Sema4A protein comprising the putative extracellular region fused to the human immunoglobulin- γ 1 (IgG1) Fc fragment (Sema4A–Fc). We generated antibodies against Sema4A by immunizing rats with Sema4A–Fc and screening the hybridomas for reactivity to mouse Chinese hamster ovary (CHO) cell transfectants expressing Sema4A (Fig. 1a). The anti-Sema4A antibody (hereafter referred to as anti-Sema4A) specifically bound to Sema4A but not to control CHO cells carrying only the Neomycin resistance plasmid or to CHO cells expressing CD100. As expected from our cloning methodology, Sema4A was expressed abundantly on the surface of bone-marrow-derived and splenic DCs (Fig. 1b). No differences were found in Sema4A expression between DCs expressing the CD8 α antigen and those not expressing it. Expression of Sema4A was also detected on the surface of resting B cells but not resting T cells. When B cells were stimulated with an antibody against CD40 (anti-CD40), the expression of Sema4A was upregulated. The slight expression of Sema4A became detectable when T cells were stimulated with an antibody against CD3 (anti-CD3).

To test whether Sema4A has an effect on T-cell activation, T cells carrying the CD4 antigen (CD4⁺ T cells) were stimulated with immobilized anti-CD3 in the presence of Sema4A–Fc or control human IgG1. Sema4A–Fc enhanced the T-cell proliferation and interleukin 2 (IL-2) production induced by anti-CD3 (Fig. 2a). We next examined whether Sema4A promotes the differentiation of naive T cells into T-helper 1 (T_H1)-like or T_H2-like effector populations under their respective culture conditions¹⁷. Addition of Sema4A–Fc significantly enhanced the induction of cells producing interferon- γ (IFN- γ) and cells producing IL-4 (Fig. 2b). We also examined the effect of Sema4A–Fc on mixed lymphocyte reactions (MLRs) between allogeneic T cells and DCs. Bone-marrow-derived DCs on a C57BL/6 background were used to stimulate BALB/c spleen CD4⁺ T cells. Sema4A–Fc enhanced MLRs, including the proliferation of T cells and the production of IL-2 (Fig. 2c). When fully mature DCs that had been treated with anti-CD40 and fixed with paraformaldehyde were used as MLR stimulants, Sema4A–Fc showed an enhancing effect (Fig. 2d), indicating that Sema4A acts directly on T cells. Anti-Sema4A also inhibited MLRs between allogeneic T cells and DCs (Fig. 2e), confirming the involvement of Sema4A in T-cell activation through interactions between T cells and DCs. Although we have shown that CD100 enhances the activation of B cells and DCs^{7,18}, Sema4A–Fc did not possess this activity (Supplementary Information Fig. 1).

To determine whether Sema4A is involved in antigen-specific T-cell responses *in vivo*, we immunized mice subcutaneously with keyhole limpet haemocyanin (KLH) in complete Freund's adjuvant (CFA) and then treated them intravenously with Sema4A–Fc for 4 d. Five days after immunization, CD4⁺ T cells were prepared from the draining lymph nodes and were tested *in vitro* for antigen-specific responses. There was a significant increase in proliferation

and in both IL-4 and IFN- γ production by CD4⁺ T cells in mice treated with Sema4A-Fc (Fig. 3a). A single injection of Sema4A-Fc on day 0 weakly enhanced the generation of antigen-specific T cells; however, delayed administration of Sema4A-Fc starting on day 1 after immunization had no effects (Supplementary Information Fig. 2). These results suggest that Sema4A-Fc functions in a relatively early phase of *in vivo* T-cell activation.

To investigate further the involvement of Sema4A in *in vivo* immune responses, we used an experimental autoimmune encephalomyelitis (EAE) model induced by the myelin oligodendrocyte glycoprotein (MOG) peptide^{19,20}. Mice immunized with the MOG peptide and treated with pertussis toxin were injected intravenously with anti-Sema4A or control rat IgGs for 4 d. The development of EAE was suppressed significantly in mice treated with anti-Sema4A as compared with control mice treated with rat IgGs (Fig. 3b). Histological analysis of the central nervous system showed marked infiltration of mononuclear inflammatory cells in the spinal cord of control mice in which EAE developed; by contrast, no such infiltration was seen in mice treated with anti-Sema4A (Supplementary Information Fig. 3). But anti-Sema4A did not have any effects

on EAE when antibody administration started on day 10 after the immunization (Supplementary Information Fig. 4), which indicates that anti-Sema4A blocks an early phase of T-cell responses against MOG.

To determine the potential mechanisms responsible for reducing the development of EAE, CD4⁺ T cells were purified from the draining lymph nodes of immunized mice and re-stimulated *in vitro* with the MOG peptide in the presence of antigen-presenting cells. Antigen-specific T-cell responses were impaired severely in mice treated with anti-Sema4A (Fig. 3c). In addition, treatment with anti-Sema4A could block the generation of antigen-specific T cells against protein antigens such as KLH (Supplementary Information Fig. 5). Although a possible mechanism for the inhibitory effect of the antibody is the deletion of Sema4A-expressing cells, administration of anti-Sema4A did not affect the population of T cells, B cells or DCs (Supplementary Information Fig. 6). Collectively, these findings indicate that anti-Sema4A inhibits the generation of antigen-specific T cells by neutralizing the effects of Sema4A, which results in inhibition of the development of EAE. This observation suggests that Sema4A has a

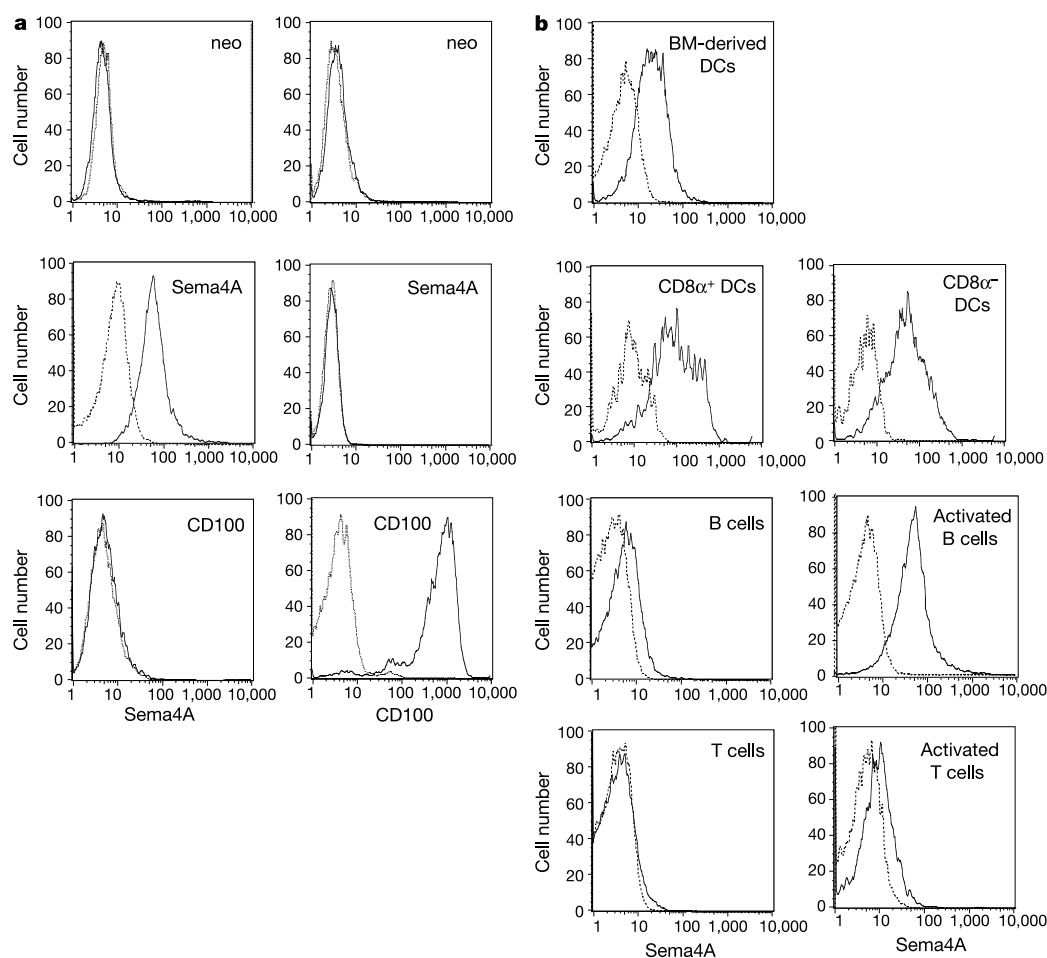


Figure 1 Analysis of Sema4A expression. **a**, Anti-Sema4A antibody specifically binds to Sema4A. CHO cells expressing Sema4A, CD100 or the Neomycin resistance gene were stained with biotinylated anti-Sema4A, anti-CD100 (unbroken lines) or biotinylated isotype-matched control (dotted lines) antibodies plus fluorescein isothiocyanate (FITC)-conjugated streptavidin. **b**, Sema4A is preferentially expressed in DCs. Bone-marrow-derived and splenic DCs were stained with FITC-conjugated anti-CD11c, phycoerythrin (PE)-conjugated anti-CD8 α and biotinylated anti-Sema4A (unbroken line) or with

biotinylated isotype-matched control (dotted line) antibodies plus allophycocyanin-conjugated streptavidin. B220-expressing and Thy-1-expressing cells prepared from the spleen were stimulated, respectively, with anti-CD40 (1 $\mu\text{g ml}^{-1}$) and immobilized anti-CD3 (5 $\mu\text{g ml}^{-1}$) for 24 h, and stained with biotinylated anti-Sema4A (unbroken line) or biotinylated isotype-matched control (dotted line) antibodies plus allophycocyanin-conjugated streptavidin before and after stimulation. The cells were analysed by flow cytometry.

crucial role in physiological and pathological cellular immune responses.

To determine the cellular expression of *Sema4A* receptors, we incubated various cells with biotinylated *Sema4A*-Fc. Binding of *Sema4A*-Fc was not detected on primary T cells, B cells or DCs (Fig. 4a), even after B cells and DCs were stimulated with anti-CD40. By contrast, stimulation of T cells with concanavalin A (ConA) induced *Sema4A*-Fc binding, which suggests that the *Sema4A* receptor is expressed upon T-cell activation. *Sema4A*-binding sites were also observed on the surface of some T-cell lines, such as EL-4

cells. In addition, the binding of *Sema4A*-Fc to activated T cells was blocked by anti-*Sema4A* (Fig. 4b). To clone the *Sema4A* receptor, we constructed an expression complementary DNA library from EL-4 cells. Plasmid DNA from the library was introduced into COS7 cells. The transfected COS7 cells were incubated with biotinylated *Sema4A*-Fc or biotinylated human immunoglobulin Fc fractions followed by streptavidin-conjugated magnetic beads. Cells that bound to *Sema4A*-Fc were enriched by magnetic sorting. A discrete band corresponding to an insert of roughly 1.0 kilobase pairs was seen after a third round of sorting (Fig. 4c), whereas no bands were apparent in cells that bound human immunoglobulin Fc fractions (data not shown). Sequencing of the cDNA insert of these clones identified the full-length cDNA encoding Tim-2, which belongs to the Tim protein family characterized by expression on T cells and conserved immunoglobulin and mucin domains²¹.

Genetic polymorphisms of Tim-1 correlate with susceptibility to airway hypersensitivity in the mouse²¹, and Tim-3 is expressed exclusively on T_H1 cells and has an essential role in T_H1 responses and macrophage activation²². By contrast, the precise functions of Tim-2 have not been elucidated. Tim-2 cDNA has been identified from a cDNA library of T cells stimulated by ConA²¹, which is in good agreement with our observation that *Sema4A*-Fc binds to ConA-stimulated, but not resting, T cells. Indeed, Tim-2 transcripts were upregulated in ConA-stimulated T cells (data not shown). *Sema4A*-Fc specifically bound to COS7 cells transfected with Tim-2 cDNA but not to COS7 cells transfected with Tim-3 cDNA or vector alone (Fig. 4d). In addition, the binding of *Sema4A* to COS7 cells

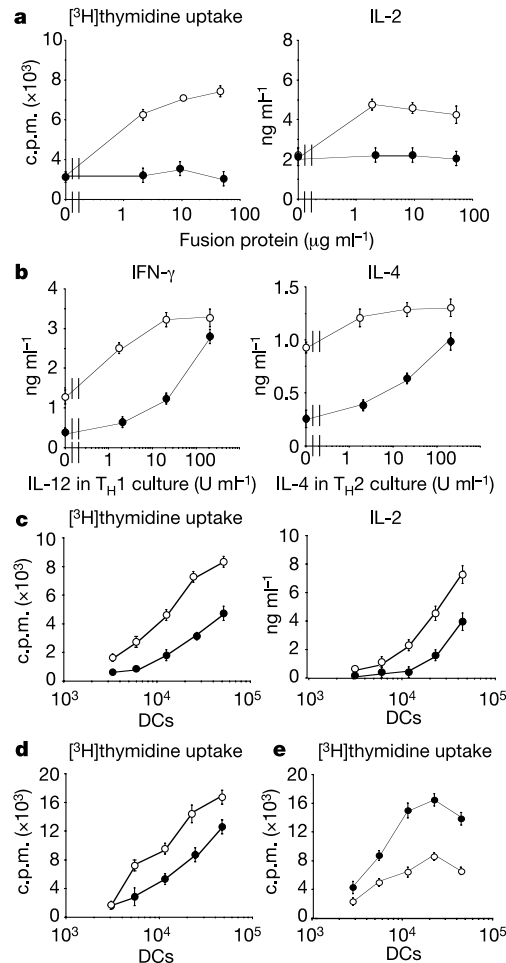


Figure 2 *Sema4A* enhances T-cell activation *in vitro*. **a**, *Sema4A*-Fc enhances T-cell activation *in vitro*. CD4⁺ T cells were stimulated with anti-CD3 with or without various concentrations of *Sema4A*-Fc (open circles) or control human IgG1 (filled circles) for 48 h. **b**, *Sema4A*-Fc promotes the differentiation of naive T cells. Naive T cells were cultured with various concentrations of IL-12 plus anti-IL-4 antibody (T_H1 conditions) or with various concentrations of IL-4 plus anti-IFN- γ antibody (T_H2 conditions) in the presence of *Sema4A*-Fc (open circles) or control human IgG1 (filled circles) for 6 d, and the collected cells were re-stimulated with anti-CD3 plus anti-CD28 for 48 h. The production of IFN- γ or IL-4 in respective T_H1 and T_H2 culture supernatants was measured by ELISA. **c**, *Sema4A*-Fc enhances MLRs. C57BL/6 DCs were incubated with BALB/c CD4⁺ T cells in the presence of *Sema4A*-Fc (open circles) or human IgG1 (filled circles) for 72 h. **d**, Effects of *Sema4A*-Fc on MLRs using fixed, fully mature DCs as stimulators. Fixed, fully mature C57BL/6 DCs were used for MLRs with BALB/c CD4⁺ T cells in the presence of *Sema4A*-Fc (open circles) or human IgG1 (filled circles). **e**, Anti-*Sema4A* antibody inhibits MLRs. Bone-marrow-derived DCs from C57BL/6 mice were stimulated with anti-CD40 for 24 h, irradiated (3,000 rad) and cultured with CD4⁺ T cells (1 $\times 10^5$ cells per well) derived from BALB/c mice with anti-*Sema4A* (20 μ g ml⁻¹, open circles) or rat IgG2a (20 μ g ml⁻¹, filled circles) for 72 h.

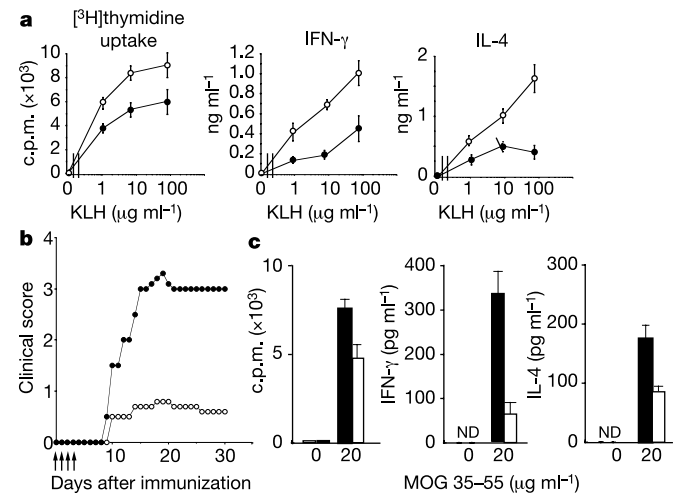


Figure 3 *Sema4A* is involved in T-cell activation *in vivo*. **a**, *Sema4A*-Fc enhances the priming of antigen-specific T-cells. C57BL/6 mice immunized with KLH received *Sema4A*-Fc (open circles) or human IgG1 (filled circles) intravenously for 4 d. Five days after immunization, CD4⁺ T cells from the draining lymph nodes were re-stimulated for 72 h with various concentrations of KLH in the presence of irradiated splenocytes from C57BL/6 mice. **b**, Clinical course of EAE^{19,20} in mice treated with anti-*Sema4A* ($n = 10$, open circles) or isotype-matched rat IgG2a ($n = 10$, filled circles). Mice were immunized with 100 μ g of MOG peptide in CFA and scored for clinical signs of EAE as described^{19,20}. Mice were treated intravenously with 100 μ g of antibodies for 4 d (arrows). The mean clinical score of each group is plotted against the time after immunization. **c**, *In vitro* responses of CD4⁺ T cells to MOG peptide. Five days after immunization, CD4⁺ T cells were prepared from the draining lymph nodes of the mice treated with anti-*Sema4A* (open bars) or isotype-matched rat IgG2a (filled bars) and stimulated for 72 h with or without the MOG peptide in the presence of irradiated syngeneic splenocytes. Cells were pulsed with 2 μ Ci of $[^3H]$ thymidine for the last 12 h. IL-4 and IFN- γ in the culture supernatants were measured by ELISA. Data are the mean $\pm$ s.e.m. of triplicate wells. ND, not detectable.

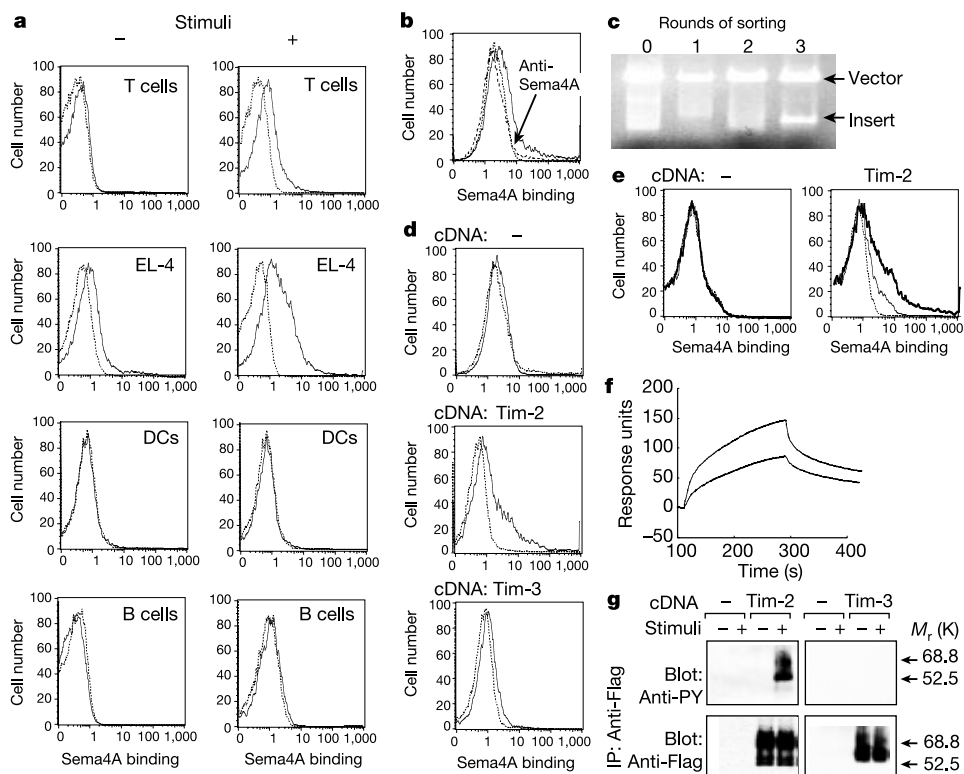


Figure 4 Tim-2 is a receptor for Sema4A. **a**, Binding of Sema4A-Fc to activated T cells. Splenic B cells, bone-marrow-derived DCs, splenic T cells or EL-4 cells were stained with biotinylated Sema4A-Fc (unbroken lines) or biotinylated human IgG1 (dotted lines) plus allophycocyanin-conjugated streptavidin before and 16 h after stimulation ($1 \mu\text{g ml}^{-1}$ anti-CD40 for B cells and DCs, and $2 \mu\text{g ml}^{-1}$ ConA for T cells and EL-4 cells) and analysed by flow cytometry. **b**, Anti-Sema4A antibody blocks the binding of Sema4A-Fc to T cells stimulated by ConA. Splenic T cells were stimulated with $2 \mu\text{g ml}^{-1}$ ConA for 16 h. Cells were stained with biotinylated Sema4A-Fc plus allophycocyanin-conjugated streptavidin in the absence (unbroken line) or presence (dotted line) of anti-Sema4A, or stained with biotinylated human IgG1 plus allophycocyanin-conjugated streptavidin (dotted line) as controls, and analysed by flow cytometry. **c**, MACS enriches a cDNA insert of about 1.0 kbp. Plasmid DNAs before (0) and after (1–3) MACS were digested with *XhoI* to identify inserts and separated on 1% agarose gels. **d**, Sema4A-Fc binds to COS7 cells that express Tim-2. COS7 cells were transfected with Tim-2 or Tim-3 cDNA, or vector alone and stained with biotinylated Sema4A-Fc (unbroken lines) or biotinylated human IgG1 (dotted lines) plus allophycocyanin-conjugated streptavidin. **e**, Tim-2-Fc blocks the

binding of Sema4A-Fc to COS7 cells transfected with Tim-2. COS7 cells were transfected with Tim-2 cDNA or vector alone and stained with biotinylated Sema4A-Fc (thick lines), biotinylated Sema4A-Fc plus an excess of Tim-2-Fc (thin line) or biotinylated human IgG1 (dotted lines) plus allophycocyanin-conjugated streptavidin. **f**, Surface plasmon resonance of Sema4A-Fc binding to Tim-2-Fc. Tim-2-Fc (100 nM, top; 50 nM, bottom) was injected through the flow cell with Sema4A-Fc immobilized on the sensor chip. As a control, Tim-2-Fc samples were injected through a flow cell without Sema4A-Fc. Binding curve represents the increase in flow cells containing Sema4A-Fc as compared with control flow cells. **g**, Sema4A-Fc induces tyrosine phosphorylation of Tim-2. COS7 cells were transfected with Flag-tagged Tim-2 or Tim-3 cDNA, or empty vector by lipofection, incubated for 2 d, and then stimulated with Sema4A-Fc ($10 \mu\text{g ml}^{-1}$). Cell lysates (generated by 1% Nonidet P-40) were immunoprecipitated with anti-Flag antibody (M2) and blotted with anti-phosphotyrosine or anti-Flag antibody. The relative molecular masses (M_r) of the marker proteins are indicated on the right.

expressing Tim-2 was specifically blocked by an excess of Tim-2-Fc (Fig. 4e). Using surface plasmon resonance, we estimated the dissociation constant (K_d) of Sema4A and Tim-2 to be $7.0 \times 10^{-8} \text{ M}$ (association rate constant, $k_{\text{on}} = 7.0 \times 10^4 \text{ M}^{-1} \text{ s}^{-1}$, dissociation rate constant $k_{\text{off}} = 4.9 \times 10^{-3} \text{ s}^{-1}$; Fig. 4f), a relatively higher value than that known for CD100–CD72 interactions ($2.0 \times 10^{-7} \text{ M}$)¹⁰.

The presence of an intracellular tyrosine kinase phosphorylation motif, Arg-Thr-Arg-Cys-Glu-Asp-Gln-Val-Tyr, in Tim-2 (ref. 21) led us to test the effect of Sema4A stimulation on its phosphorylation state. Sema4A stimulation induced the tyrosine phosphorylation of Tim-2 but not Tim-3 in COS7 cells 5 min after stimulation (Fig. 4g), suggesting that Tim-2 is a functional receptor that may transduce signals through the phosphorylated tyrosine residue. This finding differed, however, from our previous observation that CD100 induces tyrosine dephosphorylation of its receptor, CD72—an inhibitory receptor that belongs to the C-type lectin

family^{10,23,24}. Our findings indicate that, although both Sema4A and CD100 belong to the class IV semaphorin subfamily, they use structurally distinct receptors and deliver biochemically distinct signals in their respective target cells. In addition to CD72, plexin B-1 has been identified as a receptor for CD100 in non-lymphoid tissues¹⁴. At present, we do not know whether Sema4A also uses a member of the plexin family.

We have shown that Sema4A, which is expressed on DCs, enhances T-cell activation and that Tim-2 is a receptor for Sema4A, although we cannot rule out the possibility that other molecules may be involved in the T-cell activation mediated by Sema4A. We showed previously that CD100 has an essential role in T-cell activation by inducing the maturation of DCs^{7,18}. Taken together, our findings show that two class IV semaphorins, Sema4A and CD100 (Sema4D), have crucial roles in the reciprocal stimulation between T cells and antigen-presenting cells. These molecules may form a family of immunoregulatory molecules and

might constitute targets for potentiation of, and intervention in, immune responses. □

Methods

Isolation of Sema4A cDNA

We used degenerate primers based on sequences conserved among members of the semaphorin family (5'-AARTGGACIACITTYTIAARGC-3') and (5'-TCCCAIGCRCA RTRIGGRTC-3') to amplify by PCR cDNA isolated from bone-marrow-derived DCs from CD100-deficient mice¹¹. The PCR conditions were 30 cycles of 94 °C for 1 min, 55 °C for 1 min, and 72 °C for 1 min. Amplification products were cloned into a T/A vector (Novagen) and sequenced. A BLAST search of the National Center for Biotechnology Information mouse expressed sequence tag (EST) database identified the mouse Sema4A cDNA. Using this sequence, we cloned the full-length cDNA from a cDNA library generated from bone-marrow-derived DCs by PCR using a forward primer containing a *Sall* restriction (5'-AGGTCGACCCATCTGGTGACCATCTCAGGCTGACCATGGC-3') and a reverse primer containing a *NotI* site and Flag tag (5'-ATGCGGCCGCTTAC TTGTCATCGTCCTTGTAGTCAGCCATTCGGCGGCC AGATGGTTG-3'). The resulting *Sall/NotI* fragments were cloned into the pEFBos vector²⁵.

Production of Sema4A-Fc and Tim-2-Fc

We prepared truncated forms of the Sema4A and Tim-2 cDNAs from their respective full-length cDNAs by PCR using a forward primer containing a *Sall* site (5'-AGGTCGACCC ATCTGGTGACCATCTCAGGCTGACCATGGC-3') and a reverse primer containing a *BglII* site (5'-ATAGATCTGTACTTACTTTGGGACGACCATGGAAGCTCCGC-3') or forward primer containing a *XhoI* site (5'-GACCTCGAGACTGGTCTGATCATGAA TCA GATCAAGTC-3') and a reverse primer containing a *BamHI* site (5'-GACGGATCCAC TTACTTTGTGGCTTCTGTGGAGGATTACTTCA-3'). The resulting *Sall/BglII* and *XhoI/BamHI* fragments were used to replace the *Sall/BamHI* DNA fragments of the pEFBos human IgG1 Fc cassette. Production of Sema4A-Fc and Tim-2-Fc has been described¹⁰.

In vitro T-cell stimulation

To assay T-cell proliferation, CD4⁺ T cells were prepared from splenocytes using magnetic cell sorting (MACS; Miltenyi Biotech). Cells (1 × 10⁵) were stimulated with anti-CD3 (2C11; 0.1 µg ml⁻¹) coated onto flat-bottomed 96-well plates for 48 h. For T-cell differentiation analysis, CD62L^{hi}CD4⁺ naive T cells purified by fluorescence-activated cell sorting (purity >98%) were stimulated with anti-CD3 (1 µg ml⁻¹) plus anti-CD28 (37.51; 10 µg ml⁻¹), with or without Sema4A-Fc (10 µg ml⁻¹) or control human IgG1 (10 µg ml⁻¹) for 6 d, supplemented with various concentrations of IL-12 plus anti-IL-4 antibody to generate T_H1-like cells or with various concentrations of IL-4 plus anti-IFN-γ antibody to generate T_H2-like cells¹⁷. The collected cells were then re-stimulated with anti-CD3 (5 µg ml⁻¹) plus anti-CD28 (10 µg ml⁻¹) antibodies for 48 h.

For MLRs, DCs were generated from the bone marrow progenitors of C57BL/6 mice, using granulocyte/macrophage colony-stimulating factor, as described²⁶. Irradiated (3,000 rad) DCs were cultured with CD4⁺ T cells (5 × 10⁴ cells per well) derived from BALB/c mice with or without Sema4A-Fc (10 µg ml⁻¹) or human IgG1 (10 µg ml⁻¹) in flat-bottomed 96-well plates for 72 h. Cells were pulsed with 2 µCi of [³H]thymidine for the last 12 h. Amounts of IL-2 in the culture supernatants were measured using enzyme-linked immunosorbent assay (ELISA; Endogen). For MLR using fixed fully mature DCs, we treated bone-marrow-derived DCs with anti-CD40 (HM40-3; 5 µg ml⁻¹) for 24 h, fixed them with 0.8% paraformaldehyde and used them as stimulators.

In vivo T-cell priming

We immunized C57BL/6 mice (8–12 weeks old) with 10 µg of KLH or 100 µg of MOG peptide in CFA into the hind footpads. Sema4A-Fc (100 µg per mouse per day) or anti-Sema4A (SK31, rat IgG2a; 100 µg per mouse per day) was injected for 4 d after immunization. Five days after immunization, CD4⁺ T cells were prepared from the draining lymph nodes by MACS and 1 × 10⁵ cells were stimulated with various concentrations of KLH or MOG peptide in the presence of irradiated (3,000 rad) syngenic splenocytes (1 × 10⁶ cells) in flat-bottomed 96-well plates for 72 h. Cells were pulsed with 2 µCi of [³H]thymidine for the last 12 h. We measured the amounts of IL-4 and IFN-γ in the culture supernatants by ELISA (R&D systems).

Expression cloning

A cDNA library for expression cloning was generated from EL-4 cells as described¹⁰. Aliquots of 2 × 10⁷ independent clones were transfected into COS7 cells using Lipofectamine Plus 2000 (Invitrogen). Three days after transfection, the cells were collected, resuspended at a concentration of 5 × 10⁶ cells per ml in PBS containing 5% fetal calf serum, 2.5 µg ml⁻¹ of Fc block (PharMingen) and 5 µg ml⁻¹ of biotinylated Sema4A-Fc or biotinylated human immunoglobulin Fc, and incubated on ice for 1 h. The cells were washed with PBS and resuspended at 5 × 10⁶ cells per ml in PBS containing Dynabeads M-280 streptavidin (Dyna). After incubation for 30 min, the cells were washed with ice-cold PBS ten times using a Magnetic Particle Concentrator (Dyna). We extracted extrachromosomal plasmid DNA from binding cells by the Hirt method²⁷. Plasmid DNA was introduced into *Escherichia coli* DH10B cells by electroporation, and then used for a second and third transfection by protoplast fusion. Magnetic sorting was repeated three times as described above.

Surface plasmon resonance

Binding experiments were carried out by surface plasmon resonance at 25 °C on a BIAcore instrument. Sema4A-Fc protein was covalently coupled by primary amine groups to the carboxymethylated dextran matrix on a research grade CM5 sensor chip (BIAcore) using the Amine Coupling kit (BIAcore), and unreactive groups on the chip were blocked by ethanolamine according to the manufacturer's instructions. The dissociation equilibrium constant is defined as $K_d = k_{off}/k_{on}$ and was estimated by BIAcore evaluation software.

Received 5 June; accepted 10 July 2002; doi:10.1038/nature01037.

1. Semaphorin Nomenclature Committee. Unified nomenclature for the semaphorins/collapsins. *Cell* **97**, 551–552 (1999).
2. Yu, H. H. & Kolodkin, A. L. Semaphorin signaling: a little less per-plexin. *Neuron* **22**, 11–14 (1999).
3. Kolodkin, A. L., Matthes, D. J. & Goodman, C. S. The semaphorin genes encode a family of transmembrane and secreted growth cone guidance molecules. *Cell* **75**, 1389–1399 (1993).
4. Luo, Y., Raible, D. & Raper, J. A. Collapsin: a protein in brain that induces the collapse and paralysis of neuronal growth cones. *Cell* **75**, 217–227 (1993).
5. Tessier-Lavigne, M. & Goodman, S. C. The molecular biology of axon guidance. *Science* **274**, 1123–1133 (1996).
6. Delaire, S., Elhabazi, A., Bensussan, A. & Bomsell, L. CD100 is a leukocyte semaphorin. *Cell Mol. Life Sci.* **54**, 1265–1276 (1998).
7. Kumanogoh, A. & Kikutani, H. The CD100-CD72 interaction: a novel mechanism of immune regulation. *Trends Immunol.* **22**, 670–676 (2001).
8. Comeau, M. R. *et al.* A poxvirus-encoded semaphorin induces cytokine production from monocytes and binds to a novel cellular semaphorin receptor, VESPR. *Immunity* **8**, 473–482 (1998).
9. Xu, X. *et al.* Human semaphorin K1 is glycosylphosphatidylinositol-linked and defines a new subfamily of viral-related semaphorins. *J. Biol. Chem.* **273**, 22428–22434 (1998).
10. Kumanogoh, A. *et al.* Identification of CD72 as a lymphocyte receptor for the class IV semaphorin CD100: a novel mechanism for regulating B cell signaling. *Immunity* **13**, 621–631 (2000).
11. Shi, W. *et al.* The class IV semaphorin CD100 plays nonredundant roles in the immune system: defective B and T cell activation in CD100-deficient mice. *Immunity* **13**, 633–642 (2000).
12. Wang, X. *et al.* Functional soluble CD100/Sema4D released from activated lymphocytes: possible role in normal and pathologic immune responses. *Blood* **97**, 3498–3504 (2001).
13. Watanabe, C. *et al.* Enhanced immune responses in transgenic mice expressing a truncated form of the lymphocyte semaphorin CD100. *J. Immunol.* **167**, 4321–4328 (2001).
14. Tamagnone, L. *et al.* Plexins are a large family of receptors for transmembrane, secreted, and GPI-anchored semaphorins in vertebrates. *Cell* **99**, 71–80 (1999).
15. Tamagnone, L. & Comoglio, P. M. Signalling by semaphorin receptors: cell guidance and beyond. *Trends Cell Biol.* **10**, 377–383 (2000).
16. Puschel, A. W., Adams, R. H. & Betz, H. Murine semaphorin D/collapsin is a member of a diverse gene family and creates domains inhibitory for axonal extension. *Neuron* **14**, 941–948 (1995).
17. Fang, D. *et al.* Dysregulation of T lymphocyte function in itchy mice: a role for Itch in TH2 differentiation. *Nature Immunol.* **3**, 281–287 (2002).
18. Kumanogoh, A. *et al.* Requirement for the lymphocyte semaphorin, CD100, in the induction of antigen-specific T cells and the maturation of dendritic cells. *J. Immunol.* **169**, 1175–1181 (2002).
19. Chang, T. T. *et al.* Studies in B7-deficient mice reveal a critical role for B7 costimulation in both induction and effector phases of experimental autoimmune encephalomyelitis. *J. Exp. Med.* **190**, 733–740 (1999).
20. Alexander, B. H. *et al.* DAP12-deficient mice fail to develop autoimmunity due to impaired antigen priming. *Immunity* **13**, 345–353 (2000).
21. McIntire, J. J. *et al.* Identification of Tapr (an airway hyperreactivity regulatory locus) and the linked Tim gene family. *Nature Immunol.* **2**, 1109–1116 (2001).
22. Monney, L. *et al.* Th1-specific cell surface protein Tim-3 regulates macrophage activation and severity of an autoimmune disease. *Nature* **415**, 536–541 (2002).
23. Adachi, T., Flaswinkel, H., Yakura, H., Reth, M. & Tsubata, T. The B cell surface protein CD72 recruits the tyrosine phosphatase SHP-1 upon tyrosine phosphorylation. *J. Immunol.* **160**, 4662–4665 (1998).
24. Pan, C., Baumgarth, N. & Parnes, J. R. CD72-deficient mice reveal nonredundant roles of CD72 in B cell development and activation. *Immunity* **11**, 495–506 (1999).
25. Suda, T. & Nagata, S. Purification and characterization of the Fas-ligand that induces apoptosis. *J. Exp. Med.* **179**, 873–879 (1994).
26. Inaba, K. *et al.* Generation of large numbers of dendritic cells from mouse bone marrow cultures supplemented with granulocyte/macrophage colony stimulating factor. *J. Exp. Med.* **176**, 1693–1702 (1992).
27. Seed, B. & Aruffo, A. Molecular cloning of the CD2 antigen, the T-cell erythrocyte receptor, by a rapid immunoselection procedure. *Proc. Natl Acad. Sci. USA* **84**, 3365–3369 (1987).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements We thank K. Kubota for secretarial assistance; E. L. Barsoumian for critically reading the manuscript; and K. Nakamura, K. Shiozaki, S. Koga and J. Fujikake for technical support. This study was supported by research grants from the Ministry of Education, Culture, Science and Technology of Japan to H.K. and A.K.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to H.K. (e-mail: kikutani@ragtime.biken.osaka-u.ac.jp).

Apoptosis initiated by Bcl-2-regulated caspase activation independently of the cytochrome c/Apaf-1/caspase-9 apoptosome

Vanessa S. Marsden*, Liam O'Connor†‡, Lorraine A. O'Reilly*, John Silke*, Donald Metcalf*, Paul G. Ekert‡, David C. S. Huang*, Francesco Cecconi§, Keisuke Kuida¶, Kevin J. Tomaselli#, Sophie Roy☆, Don W. Nicholson☆, David L. Vaux*, Philippe Bouillet*, Jerry M. Adams*†† & Andreas Strasser*††

* The Walter and Eliza Hall Institute, Melbourne, 3050, Australia

‡ Murdoch Children's Research Institute, Melbourne, 3050, Australia

§ Department of Biology, Università Tor Vergata, Rome, Italy

¶ Genomic Pharmacology, Vertex Pharmaceuticals, Cambridge, Massachusetts 02139, USA

Idun Pharmaceuticals, San Diego, California 92121, USA

☆ Merck-Frosst, Pointe-Claire-Dorval H9H 3L1, Canada

†† These authors contributed equally to this work

Apoptosis is an evolutionarily conserved cell suicide process executed by cysteine proteases (caspases) and regulated by the opposing factions of the Bcl-2 protein family^{1,2}. Mammalian caspase-9 and its activator Apaf-1 were thought to be essential, because mice lacking either of them display neuronal hyperplasia and their lymphocytes and fibroblasts seem resistant to certain apoptotic stimuli³⁻⁶. Because Apaf-1 requires cytochrome *c* to activate caspase-9, and Bcl-2 prevents mitochondrial cytochrome *c* release, Bcl-2 is widely believed to inhibit apoptosis by safeguarding mitochondrial membrane integrity⁷⁻⁹. Our results suggest a different, broader role, because Bcl-2 overexpression increased lymphocyte numbers in mice and inhibited many apoptotic stimuli, but the absence of Apaf-1 or caspase-9 did not. Caspase activity was still discernible in cells lacking Apaf-1 or caspase-9, and a potent caspase antagonist both inhibited apoptosis and retarded cytochrome *c* release. We conclude that Bcl-2 regulates a caspase activation programme independently of the cytochrome *c*/Apaf-1/caspase-9 'apoptosome', which seems to amplify rather than initiate the caspase cascade.

How Bcl-2 prevents apoptosis remains hotly debated^{12,7-10}. Its *Caenorhabditis elegans* homologue CED-9 binds to the adapter CED-4 to prevent it from activating the caspase CED-3 (ref. 1). Because the mammalian adaptor Apaf-1 does not bind to Bcl-2 (ref. 11) and requires cytochrome *c* to activate caspase-9, it is widely believed that the Bcl-2 family regulates mitochondrial membrane pores that release cytochrome *c* and other apoptogenic factors⁷⁻⁹. However, because mammalian Bcl-2 can inhibit cell death in *C. elegans*^{1,12}, where no evidence for mitochondrial release of apoptogenic factors has emerged¹, we and others have hypothesized that Bcl-2 can regulate activation of initiator caspases independently of the cytochrome *c*/Apaf-1/caspase-9 'apoptosome'^{2,10,11}.

To investigate whether the apoptosome is required for all apoptosis regulated by the Bcl-2 protein family, we have compared the impact of loss of Apaf-1 or caspase-9 with that caused by Bcl-2 overexpression or loss of its antagonist Bim on programmed cell death in the haematopoietic system, where Apaf-1 and caspase-9 are widely expressed (Supplementary Fig. 1). In cytokine-supported cultures, fetal liver cells from wild-type, Apaf-1^{-/-}, caspase-9^{-/-} and *vav-bcl-2* transgenic mice produced colonies of similar number, size and composition (Fig. 1a and data not shown). Without cytokine support, Bcl-2 overexpression protected the colony-

forming cells up to 100-fold, but progenitors lacking caspase-9 or Apaf-1 died as rapidly as those from wild-type mice (Fig. 1a).

Because most Apaf-1-deficient or caspase-9-deficient mice die before birth³⁻⁶, we assessed the role of the apoptosome in haematopoietic homeostasis by reconstituting this compartment of C57BL/6-Ly5.1 mice with fetal liver stem cells from the mutant embryos, or from wild-type, Bim-deficient or *vav-bcl-2* embryos. Lymphocytes of Apaf-1^{-/-} or caspase-9^{-/-} origin were no more numerous than those of wild-type origin in the thymus (Fig. 1b), spleen (Fig. 1c), lymph nodes or bone marrow (data not shown). In contrast, mice reconstituted with Bim-deficient or *vav-bcl-2* stem cells had 3- to 5-fold more splenic B and T cells (Fig. 1c), as do unmanipulated *bim*^{-/-} and *vav-bcl-2* mice^{13,14}. Thus, unlike Bim and Bcl-2, caspase-9 and Apaf-1 are not critical determinants of programmed cell death in haematopoiesis.

To compare the role of these molecules in stress-induced apoptosis, we isolated donor-derived CD4⁺8⁺ thymocytes, pro-B cells and mature T and B cells from the reconstituted mice and tested their sensitivity to cytokine withdrawal, γ -radiation, etoposide and dexamethasone. None of these death responses required Apaf-1 or caspase-9. In thymocytes and activated B and T lymphoblasts, absence of either cell death protein afforded at most modest (<2-fold) protection, whereas in mature T cells and pro-B cells, lack of caspase-9 or Apaf-1 had no impact (Fig. 2a, b, Supplementary Fig. 3 and data not shown). In contrast, Bcl-2 markedly protected against all these apoptotic stimuli, and, as reported¹³, Bim was required for death provoked by cytokine withdrawal but was largely dispensable for that induced by dexamethasone or etoposide. These findings were not restricted to haematopoietic cell types, because Apaf-1-deficient and caspase-9-deficient embryonic fibroblasts were as sensitive as wild-type cells to loss of attachment (anoikis) (Fig. 2c).

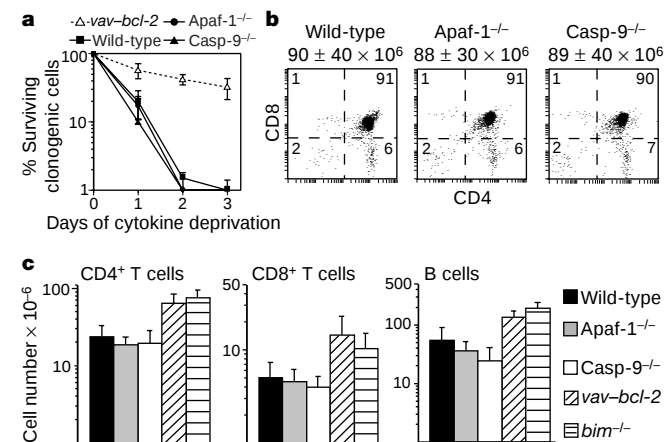


Figure 1 Cytokine withdrawal-induced and programmed death of haematopoietic cells lacking Apaf-1 or caspase-9. **a**, Fetal liver cells from embryos at embryonic day 14.5 (E14.5) (wild-type, Apaf-1^{-/-}, caspase-9^{-/-} and *vav-bcl-2* transgenic) were cultured in semi-solid agar and haematopoietic growth factors added immediately or after 1, 2 or 3 days. Colony numbers, scored after seven days of stimulation, are arithmetic means $\pm$ s.d. for 3–6 mice of each genotype. **b**, **c**, Cell numbers in lethally irradiated (2×5.5 Gy) C57BL/6-Ly5.1 mice repopulated with fetal liver cells from C57BL/6-Ly5.2 wild-type, Apaf-1^{-/-}, caspase-9^{-/-}, *bim*^{-/-} or *vav-bcl-2* transgenic mice. **b**, Thymocytes from recipient mice repopulated with wild-type, Apaf-1^{-/-} or caspase-9^{-/-} stem cells were stained with monoclonal antibodies to CD4, CD8 and Ly5.1 (profiles gated for donor-derived Ly5.1⁺ cells). Numbers at top indicate the mean thymic cellularities for 3–6 mice of each genotype $\pm$ s.d.; the percentages in the four subpopulations are indicated. (Cellularity refers to the total cell number in an organ.) **c**, Total numbers of lymphoid cell subtypes in spleens of mice reconstituted with wild-type, caspase-9^{-/-}, Apaf-1^{-/-}, *bim*^{-/-} or *vav-bcl-2* cells. Data derive from 3–6 mice of each genotype and error bars represent s.d.

† Present address: Incyte Genomics, Palo Alto, California 94304, USA

The absence of Apaf-1 or caspase-9 has previously been reported to delay the death of lymphocytes and fibroblasts induced by DNA damage or cytotoxic drugs^{3–5}, but those lymphocytes came from the few mutant mice that survived to birth, and ill health may have selected for stress-resistance. Whereas our fibroblast experiments used a physiological death stimulus, the strong cytotoxic agents used previously^{4,5} may have triggered apoptosome activation by directly damaging mitochondria. Also, some discrepancies may reflect the mixed genetic background used previously, because this affects the Apaf-1^{-/-} phenotype (ref. 15 and F.C., unpublished results).

Although Apaf-1-deficient embryonic stem cells and fibroblasts treated with cytotoxic agents were reported to die by a non-apoptotic mechanism^{16,17}, we found that the mutant cells underwent apoptosis. Mutant T lymphoblasts deprived of cytokine exhibited condensed chromatin and plasma membrane blebbing (Fig. 3a), and dying thymocytes and fibroblasts exhibited hallmarks of apoptosis, including cell-surface exposure of phosphatidylserine (Supplementary Fig. 4c, d), loss of mitochondrial membrane potential (Supplementary Fig. 4a) and inter-nucleosomal DNA cleavage, as revealed by DNA laddering (Fig. 3b) and the TUNEL reaction (TdT-mediated dUTP nick end labelling; see Supplementary Fig. 4b).

Are caspases activated in the absence of the apoptosome? In the γ -irradiated mutant cells, the two classical caspase substrates PARP (poly(ADP-ribose) polymerase) and ICAD (inhibitor of the caspase-activated DNase, also known as DFF45; refs 18, 19) yielded fragments of the sizes seen in wild-type cells, albeit less efficiently (Fig. 4a). The ICAD cleavage is consistent with the inter-nucleo-

mal DNA degradation (Fig. 3c), because ICAD proteolysis is necessary to free CAD, the relevant DNase^{18,19}. PARP and ICAD are substrates of the closely related effector caspases-3 and -7, which have a preferred DEVD specificity^{18,19}, and lysates of the dying cells exhibited DEVDase activity with a fluorogenic substrate, albeit about tenfold less than in the wild-type cells (Fig. 4b). This activity was due to caspases, because the inhibitor DEVD-CHO blocked it (Fig. 4b). To determine which effector caspases were activated, immunoblots were examined for processing of the zymogens (inactive precursors). In γ -irradiated mutant thymocytes, no processing of pro-caspase-3 or -6 was seen (Supplementary Fig. 5). The relevant effector caspase is probably caspase-7, because its precursor was processed discernibly, albeit about tenfold less than in wild-type cells (Fig. 4a), and because extracts from dying caspase-9^{-/-} cells cleaved wild-type ICAD but not ICAD with a mutation (D224E) in the known caspase-7 or -3 recognition site (DAVD₂₂₄)^{18,19} (Fig. 4c). The ICAD cleavage is linked to apoptosis, because none was detectable in irradiated cells protected by Bcl-2 overexpression (Fig. 5c).

To determine whether caspase activity caused the death of the mutant cells, we tested two recently described peptido-mimetic inhibitors that block multiple caspases with high specificity (refs 20, 21 and Supplementary Table 1). IDN-1965 strongly inhibited dexamethasone-induced apoptosis in wild-type as well as mutant T cells (Fig. 5a). It blocked apoptosis and ICAD cleavage nearly as well as Bcl-2 overexpression and did not further enhance survival of

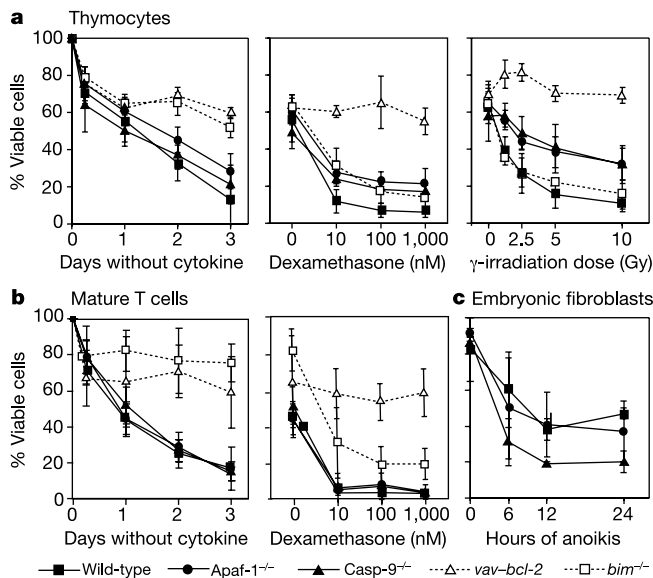


Figure 2 Caspase-9 and Apaf-1 are largely dispensable for death induced by growth-factor withdrawal- and stress-induced death of lymphocytes and fibroblasts. **a, b**, FACS sorting was used to purify CD4⁺8⁺ (donor-derived Ly5.1⁻) thymocytes (**a**) and mature T cells (**b**) from lethally irradiated C57BL/6-Ly5.1 mice reconstituted with wild-type, Apaf-1^{-/-}, caspase-9^{-/-}, *bim*^{-/-} or *vav-bcl-2* fetal liver cells (C57BL/6-Ly5.2). Cells were cultured in simple medium (no growth factor) over 3 days, or for 1 day in the presence of graded doses of dexamethasone or after γ -irradiation. Cell viability was determined by propidium iodide staining and flow cytometric analysis. **c**, Early passage wild-type, Apaf-1^{-/-} and caspase-9^{-/-} embryonic fibroblasts were cultured for 24 h under conditions preventing cell attachment (poly-haema-coated bacterial culture-ware). Apoptosis was assessed by FACS to determine the proportion of cells with a hypo-diploid DNA content (see Methods in the Supplementary Information). Data represent arithmetic means $\pm$ s.d. of 3–6 independent experiments with cells from each genotype.

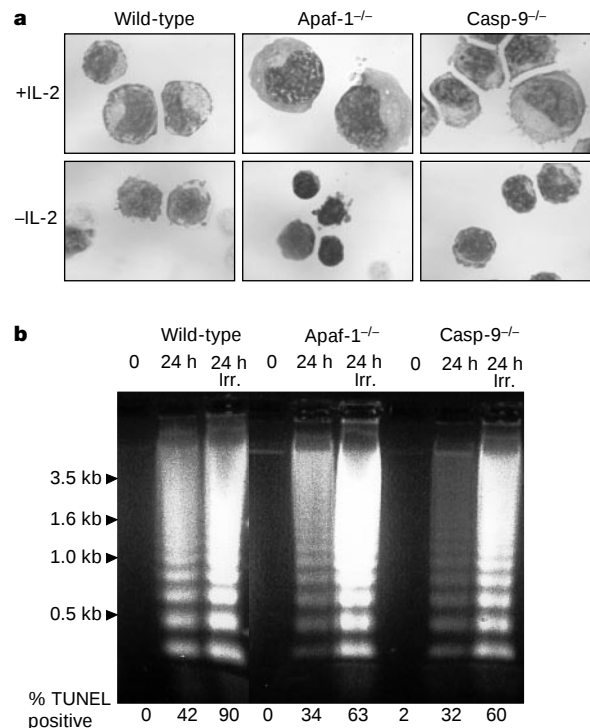


Figure 3 Apoptotic hallmarks in dying caspase-9-deficient and Apaf-1-deficient lymphocytes. **a**, Activated T cells were cultured for 24 h in the presence or absence of interleukin-2 ($\pm$ IL-2). Cytospins were stained as described in the Supplementary Methods and viewed at 500 $\times$ magnification. **b**, Thymocytes were sampled either initially (0 h) or 24 h after culture in medium alone, either untreated (spontaneous death) or after 5 Gy γ -irradiation (Irr.). Demonstration of DNA degradation typical of apoptosis by 'laddering' of nuclear DNA electrophoresed on an agarose gel and by the TUNEL reaction (percentage of cells with incorporation of FITC-labelled dUTP indicated; see Supplementary Fig. 4b for TUNEL profiles measured by FACS). TUNEL assay, TdT-mediated dUTP nick end labelling.

the *bcl-2* transgenic cells (Fig. 5). The structurally distinct inhibitor IDN-6275 (refs 20, 21) was as effective (data not shown), and the widely used peptide-based caspase inhibitor zVAD-fmk also inhibited apoptosis, although less well, presumably owing to its limited membrane permeability and short half-life. Notably, in γ -irradiated wild-type and Apaf-1-deficient thymocytes, IDN-1965 also retarded cytochrome *c* release from mitochondria (Fig. 6a). Thus, caspase activity may be required for mitochondrial membrane disruption.

To purify the active caspases responsible for apoptosis, cell extracts were incubated with the irreversible pseudo-substrate biotin-DEVD-aomk (at a concentration high enough to react with most caspases) and biotinylated polypeptides were recovered on a streptavidin column²². Blots of polypeptides from immortalized Apaf-1^{-/-} myeloid cells treated with etoposide yielded at

least three polypeptides (p18, p26 and p35) absent from untreated cells (Fig. 6b). In accord with Fig. 4, immunoblotting identified the p18 fragment as the large subunit of caspase-7, whereas the p26 fragment represented processed caspase-1 (Fig. 6b). Although the p35 fragment did not react with antibodies to caspases 1, 2, 3, 6, 7, 8 or 9, it has the size expected for a partially processed initiator caspase such as caspase-11 or -12.

Our results establish that the cell-death pathway controlled by Bcl-2 does not require caspase-9 or its activator Apaf-1. In keeping with our evidence (Fig. 1b, c) that neither is required for haematopoietic homeostasis, in which the Bcl-2 family has major roles¹⁰, deletion of thymocytes with self-reactivity depends on Bim²³ but not on Apaf-1 (ref. 24). Because apoptosis was at most only slightly delayed by the absence of Apaf-1 or caspase-9 (Figs 2, 3 and Supplementary Figs 3, 4), we conclude that the apoptosome is not an essential trigger for apoptosis but is rather a machine for amplifying the caspase cascade (Fig. 6c). Presumably this amplification is needed much more in some cell types (for example, neuronal precursors) than others (for example, lymphocytes), perhaps depending on their complement of caspases, level of caspase inhibitors (for example, IAPs) or content of vital substrates. Although mitochondrial release of apoptogenic molecules contributes to cell destruction⁷⁻⁹, apoptosis in the absence of the

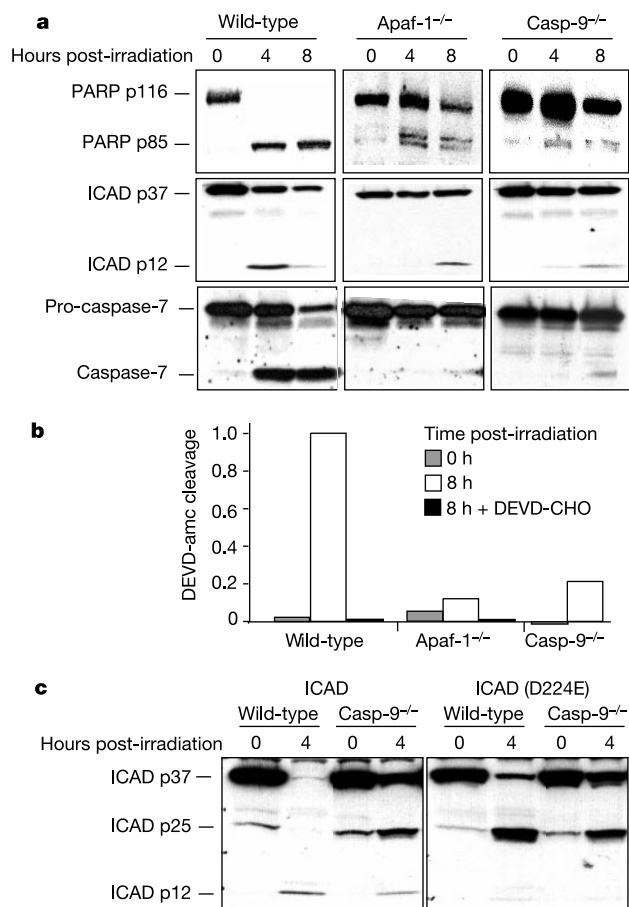


Figure 4 Caspase activity and cleavage of caspase substrates in absence of Apaf-1 or caspase-9. **a**, Thymocytes were cultured for 0, 4 and 8 h following 5 Gy γ -irradiation. Cell extracts were western blotted with antibodies specific for PARP, ICAD and caspase-7 (Supplementary Methods). **b**, Residual DEVD-amc cleaving activity in the absence of Apaf-1 or caspase-9. Lysates from thymocytes were made immediately after 5 Gy γ -irradiation or after 8 h culture and assayed for ability to cleave DEVD-amc to release a fluorescent product. The cleavage was shown to reflect caspase activity by incubation with 200 nM DEVD-CHO (see Supplementary Methods). Data, expressed as relative rates of degradation of DEVD-amc over 40 min at 37 °C, are representative of several independent experiments. **c**, ICAD is processed at a site cleaved only by caspases 3 and 7. Thymocyte lysates were prepared and incubated with exogenous wild-type ICAD or ICAD having a mutation (D224E) that abolishes the caspase-3 or -7 processing site^{18,19}, synthesized *in vitro* as described in the Supplementary Methods. Proteins were then separated by SDS-polyacrylamide gel electrophoresis (PAGE) and western blotted to detect ICAD processing. The p25 fragment is generated by cleavage at D117; the p12 at D224 (refs 18, 19). DEVD-amc, Asp-Glu-Val-Asp-aminomethylcoumarin; DEVD-CHO, Asp-Glu-Val-Asp-aldehyde.

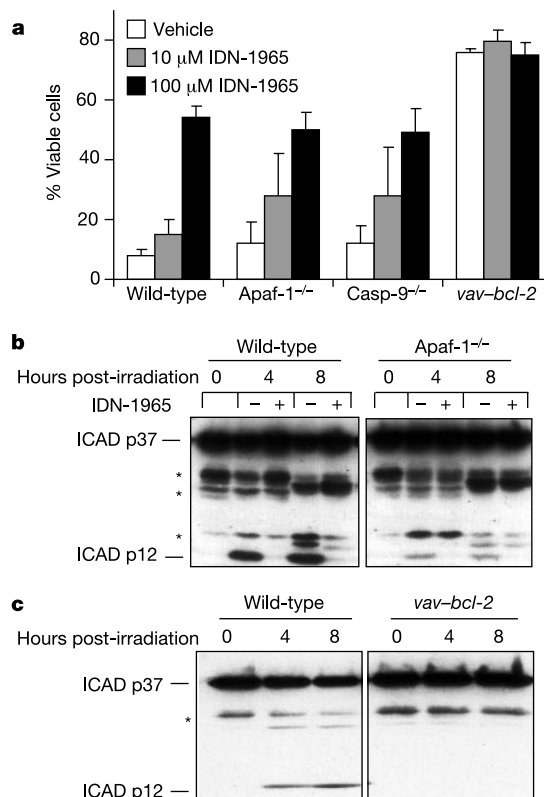


Figure 5 Death of Apaf-1-deficient and caspase-9-deficient cells and ICAD proteolysis is caspase-dependent. **a**, Wild-type, Apaf-1^{-/-}, caspase-9^{-/-} and *vav-bcl-2* T cells were treated with 10⁻⁷ M dexamethasone for 24 h in the presence of 10 or 100 μ M IDN-1965 (ref. 20) or vehicle (0.4% DMSO) alone. Cell viability (mean $\pm$ s.d. of 2–6 independent experiments) was assessed by propidium iodide staining of unfixed cells and FACS analysis. IDN-1965 protected cultured cells against apoptosis for at least 3 days. **b**, Wild-type and Apaf-1^{-/-} thymocytes were cultured following 5 Gy γ -irradiation for 0, 4 or 8 h in the presence (+) or absence (–) of 100 μ M IDN-1965, and cell extracts western blotted with an antibody specific for ICAD. **c**, Wild-type and *vav-bcl-2* thymocytes were cultured following 5 Gy γ -irradiation for 0, 4 or 8 h, and cell extracts western blotted with an antibody specific for ICAD. Asterisks indicate non-ICAD proteins cross-reactive with the immunoblotting reagents.

apoptosome seems to depend on caspase activity (Fig. 5a) and caspase-7 was identified as the probable effector (Figs 4, 6b). The low level of caspase-7 activity observed (Figs 4a, b) would be expected in the absence of amplification by the apoptosome^{3–6} but appears to be sufficient to demolish many cell types. Although all caspase activation has been thought to require mitochondrial disruption^{7–9}, efficient methods of caspase inhibition (Fig. 5 and ref. 25) reveal that caspase activation may instead be required for mitochondrial membrane disruption. Cell death in *Drosophila* also seems not to require cytochrome *c* release^{26,27}.

Which upstream caspase(s) might be responsible? In certain transformed cells caspase-2 has been implicated²⁵, but it cannot be the sole initiator, because the caspase-2 knockout phenotype is unremarkable²⁸ and caspase-2^{−/−} lymphocytes and neurons retain normal sensitivity to apoptotic stimuli²⁹. In dying Apaf-1-deficient cells, we identified active caspase-1 and a polypeptide that may be processed caspase-11 or -12 (Fig. 6b). Because mice lacking caspases 1, 2, 11 or 12 display no gross defects in apoptosis¹⁰, we propose that developmental and stress-induced apoptosis can be triggered by several initiator caspases acting in a redundant manner and that, like CED-9 in *C. elegans*, Bcl-2 controls their activators (Fig. 6c). These initiator caspases could either directly activate caspase-7, bypassing the mitochondrion, or mediate disruption of its outer membrane, to amplify the caspase cascade through the apoptosome.

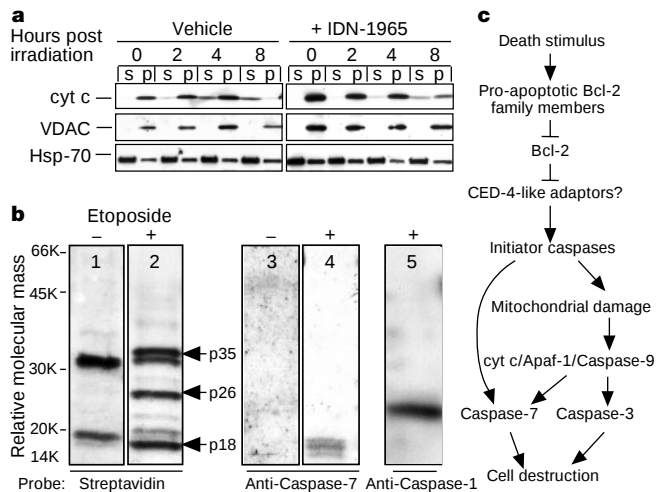


Figure 6 Role of caspases and mitochondrial disruption in apoptosis. **a**, Cytochrome *c* release may be caspase dependent. Wild-type thymocytes were cultured following 5 Gy γ -irradiation for 0, 2, 4 or 8 h in the presence or absence of 100 μ M IDN-1965. Cell lysates were fractionated (Supplementary Methods) into cytosolic proteins (s) and pelleted organelles (p). Fractions were western blotted with an antibody to cytochrome *c*. Reprobing with an antibody to VDAC (voltage-dependent anion channel) demonstrates the effectiveness of the fractionation, whereas Hsp-70 acts as a loading control for the cytosolic fractions. Similar results were obtained with Apaf-1^{−/−} thymocytes. **b**, Immortalized Apaf-1^{−/−} pro-myeloid cells were cultured either untreated (−) or in the presence of 10 μ g ml^{−1} etoposide (+). Lysates were reacted with 10 μ M biotin-DEVDA-omk (see text), and the biotinylated polypeptides were purified on a streptavidin column, and recovered by prolonged elution with 25 mM biotin. Eluted material was resolved by SDS-PAGE and blotted with horseradish-peroxidase-conjugated streptavidin to reveal biotinylated proteins (panels 1 and 2), and with antibodies to caspases 7 (panels 3 and 4) or 1 (panel 5). Arrows indicate the molecular weight of polypeptides specific to dying Apaf-1^{−/−} cells. Lane 3 was deliberately overexposed to demonstrate that no active caspase-7 was recovered in healthy cells. **c**, Model for the control of apoptosis. Bcl-2 is proposed to inhibit the activity of presumptive CED-4-like adaptors for at least two upstream initiator caspases. A few candidates for such adaptors have been described recently³⁰. The upstream initiator caspases may then directly or indirectly disrupt the outer mitochondrial membrane, activating the apoptosome, or instead directly activate caspase-7. Bax or Bak is required, but their position is unclear².

Received 4 July; accepted 3 September 2002; doi:10.1038/nature01101.
Published online 25 September 2002.

- Hengartner, M. O. The biochemistry of apoptosis. *Nature* **407**, 770–776 (2000).
- Cory, S. & Adams, J. M. The Bcl-2 family: Regulators of the cellular life-or-death switch. *Nature Rev. Cancer* **2**, 647–656 (2002).
- Kuida, K. *et al.* Reduced apoptosis and cytochrome *c*-mediated caspase activation in mice lacking caspase 9. *Cell* **94**, 325–337 (1998).
- Hakem, R. *et al.* Differential requirement for caspase 9 in apoptotic pathways in vivo. *Cell* **94**, 339–352 (1998).
- Yoshida, H. *et al.* Apaf1 is required for mitochondrial pathways of apoptosis and brain development. *Cell* **94**, 739–750 (1998).
- Cecconi, F., Alvarez-Bolado, G., Meyer, B. L., Roth, K. A. & Gruss, P. Apaf-1 (CED-4 homologue) regulates programmed cell death in mammalian development. *Cell* **94**, 727–737 (1998).
- Green, D. R. & Reed, J. C. Mitochondria and apoptosis. *Science* **281**, 1309–1311 (1998).
- Gross, A., McDonnell, J. M. & Korsmeyer, S. J. Bcl-2 family members and the mitochondria in apoptosis. *Genes Dev.* **13**, 1899–1911 (1999).
- Wang, X. The expanding role of mitochondria in apoptosis. *Genes Dev.* **15**, 2922–2933 (2001).
- Strasser, A., O'Connor, L. & Dixit, V. M. Apoptosis signaling. *Annu. Rev. Biochem.* **69**, 217–245 (2000).
- Hausmann, G. *et al.* Pro-apoptotic apoptosis protease-activating factor 1 (Apaf-1) has a cytoplasmic localization distinct from Bcl-2 or Bcl-x_L. *J. Cell Biol.* **149**, 623–634 (2000).
- Vaux, D. L., Weissman, I. L. & Kim, S. K. Prevention of programmed cell death in *Caenorhabditis elegans* by human bcl-2. *Science* **258**, 1955–1957 (1992).
- Bouillet, P. *et al.* Proapoptotic Bcl-2 relative Bim required for certain apoptotic responses, leukocyte homeostasis, and to preclude autoimmunity. *Science* **286**, 1735–1738 (1999).
- Ogilvy, S. *et al.* Constitutive bcl-2 expression throughout the hematopoietic compartment affects multiple lineages and enhances progenitor cell survival. *Proc. Natl Acad. Sci. USA* **96**, 14943–14948 (1999).
- Honarpour, N. *et al.* Adult Apaf-1-deficient mice exhibit male infertility. *Dev. Biol.* **218**, 248–258 (2000).
- Haraguchi, M. *et al.* Apoptotic protease activating factor 1 (Apaf-1)-independent cell death suppression by Bcl-2. *J. Exp. Med.* **191**, 1709–1720 (2000).
- Susin, S. A. *et al.* Two distinct pathways leading to nuclear apoptosis. *J. Exp. Med.* **192**, 571–580 (2000).
- Liu, X., Zou, H., Slaughter, C. & Wang, X. DFF, a heterodimeric protein that functions downstream of caspase-3 to trigger DNA fragmentation during apoptosis. *Cell* **89**, 175–184 (1997).
- Enari, M. *et al.* A caspase-activated DNase that degrades DNA during apoptosis, and its inhibitor ICAD. *Nature* **391**, 43–50 (1998).
- Wu, J. C. & Fritz, L. C. Irreversible caspase inhibitors: tools for studying apoptosis. *Methods Enzymol.* **17**, 320–328 (1999).
- Hoglen, N. C. *et al.* Characterization of the caspase inhibitor IDN-1965 in a model of apoptosis-associated liver injury. *J. Pharmacol. Exp. Therapeut.* **297**, 811–818 (2001).
- Nicholson, D. W. *et al.* Identification and inhibition of the ICE/CED-3 protease necessary for mammalian apoptosis. *Nature* **376**, 37–43 (1995).
- Bouillet, P. *et al.* BH3-only Bcl-2 family member Bim is required for apoptosis of autoreactive thymocytes. *Nature* **415**, 922–926 (2002).
- Hara, H. *et al.* The apoptotic protease-activating factor 1-mediated pathway of apoptosis is dispensable for negative selection of thymocytes. *J. Immunol.* **168**, 2288–2295 (2002).
- Lassus, P., Opitz-Araya, X. & Lazebnik, Y. Requirement for caspase-2 in stress-induced apoptosis before mitochondrial permeabilization. *Science* **297**, 1352–1354 (2002).
- Dorsten, L. *et al.* The role of cytochrome *c* in caspase activation in *Drosophila melanogaster* cells. *J. Cell Biol.* **156**, 1089–1098 (2002).
- Zimmermann, K. C., Ricci, J. E., Droin, N. M. & Green, D. R. The role of ARK in stress-induced apoptosis in *Drosophila* cells. *J. Cell Biol.* **156**, 1077–1087 (2002).
- Bergeron, L. *et al.* Defects in regulation of apoptosis in caspase-2-deficient mice. *Genes Dev.* **12**, 1304–1314 (1998).
- O'Reilly, L. A. *et al.* Caspase-2 is not required for thymocyte or neuronal apoptosis even though cleavage of caspase-2 is dependent on both Apaf-1 and caspase-9. *Cell Death Differ.* **9**, 832–841 (2002).
- Adams, J. M. & Cory, S. Apoptosomes: engines for caspase activation. *Curr. Opin. Cell Biol.* (in the press).

Supplementary Information accompanies the paper on Nature's website
(<http://www.nature.com/nature>).

Acknowledgements We thank Y. Lazebnik and P. Lassus for sharing unpublished results. We also thank P. Gruss for Apaf-1^{+/−} mice, Y. Lazebnik and X. Opitz-Araya for monoclonal antibodies to caspases 3, 7 and 9, P. Vandenabeele and M. Kalai for the anti-mouse caspase-1 antibody, R. Anderson for the anti-HSP70 antibody and S. Nagata for the ICAD constructs. We thank E. Loza, A. Milligan, C. Tilbrook, A. Naughton and J. Merryful for animal care, F. Battye, D. Kaminaris, V. Lapatis, J. Chan and C. Tarlinton for cell sorting, S. Mifsud, L. DiRago, L.-C. Zhang and L. Tai for expert technical help and G. Filby for editorial assistance. We are grateful to S. Cory, A. Harris, K. Newton and H. Puthalakath for discussions and critical reading of this manuscript. This work was supported by fellowships and grants from the Dr Josef Steiner Cancer Research Foundation, the NHMRC, the Leukemia and Lymphoma Society (SCOR Center), the Anti-Cancer Council of Victoria, the Sylvia and Charles Viertel Charitable Foundation, the NIH, the AIRC and the Commonwealth Department of Education, Science and Training. F.C. is an Assistant Telethon Scientist.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.S.
(e-mail: strasser@wehi.edu.au).

Initiation and re-initiation of DNA unwinding by the *Escherichia coli* Rep helicase

TaeKjip Ha*, Ivan Rasnik*, Wei Cheng†, Hazen P. Babcock‡, George H. Gauss†, Timothy M. Lohman† & Steven Chu‡

* Department of Physics, University of Illinois, Urbana, Illinois 61801, USA

† Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St Louis, Missouri 63110, USA

‡ Department of Physics, Stanford University, Stanford, California 94305, USA

Helicases are motor proteins that couple conformational changes induced by ATP binding and hydrolysis with unwinding of duplex nucleic acid^{1–3}, and are involved in several human diseases. Some function as hexameric rings⁴, but the functional form of non-hexameric helicases has been debated^{5–10}. Here we use a combination of a surface immobilization scheme and single-molecule fluorescence assays—which do not interfere with biological activity—to probe DNA unwinding by the *Escherichia coli* Rep helicase. Our studies indicate that a Rep monomer uses ATP hydrolysis to move toward the junction between single-stranded and double-stranded DNA but then displays conformational fluctuations that do not lead to DNA unwinding. DNA unwinding initiates only if a functional helicase is formed via additional protein binding. Partial dissociation of the functional complex during unwinding results in interruptions ('stalls') that lead either to duplex rewinding upon complete dissociation of the complex, or to re-initiation of unwinding upon re-formation of the functional helicase. These results suggest that the low unwinding processivity observed *in vitro* for Rep is due to the relative instability of the functional complex. We expect that these techniques will be useful for dynamic studies of other helicases and protein–DNA interactions.

Two single-molecule approaches have been used to visualize DNA unwinding by the highly processive *E. coli* RecBCD helicase/nuclease^{11,12} with resolutions up to 100 base pairs (bp)¹². However, many helicases display only limited processivity *in vitro*, so methods with higher base-pair resolution are desired. We have used the single-molecule fluorescence resonance energy transfer (FRET) technique^{13–16} to measure structural changes of DNA upon the action of *E. coli* Rep helicase. Not only could we detect DNA unwinding with ≤ 10 -bp resolution, but we also were able to observe events that are difficult to examine in bulk solution, such as stalls, rewinding of the DNA duplex and re-initiation of unwinding by a stalled enzyme–DNA complex.

To probe the initiation of DNA unwinding with maximum sensitivity we used an 18-bp or 40-bp duplex with a 3'-(dT)₂₀ tail with a donor (Cy3) and an acceptor (Cy5) fluorophore attached to the junction between single-stranded DNA (ssDNA) and double-stranded DNA (dsDNA) (DNA I and II, Fig. 1i). DNA molecules were specifically immobilized to a polymer-coated surface¹⁷, and were imaged using an evanescent field microscope¹⁵. We then rapidly (< 0.2 s) delivered a solution containing both Rep and ATP, and recorded the time records of unwinding via FRET (Fig. 1i). Before unwinding begins, the distance, R , between the two dyes is small ($E_{\text{FRET}} \approx 1$; $E_{\text{FRET}} \equiv 1/[1 + I_D/I_A]$ is an approximation for FRET efficiency, $1/([1 + (R/R_0)^6])$, where I_D and I_A are intensities for donor and acceptor respectively and $R_0 \approx 6$ nm). As the DNA is unwound, the time-averaged R increases (E_{FRET} decreases). If the DNA is fully unwound, the donor strand diffuses away from the surface and the fluorescence signal abruptly disappears, clearly marking the completion of unwinding. In contrast, dissociation of the helicase during unwinding can lead to rapid DNA rewinding

($E_{\text{FRET}} \rightarrow \sim 1$). The elimination of only one component among ATP, Rep and Mg^{2+} or the replacement of ATP by its poorly hydrolysed analogues (adenosine 5'-[γ -thio]triphosphate or adenosine 5'-(β , γ -imido)triphosphate) abolished all unwinding signals.

After adding Rep protein, which is a monomer in the absence of DNA¹⁸, and ATP, a characteristic time elapsed before unwinding was initiated. Its average, τ_{UI} , is inversely proportional to [Rep] (above 20 nM) and the rate of unwinding initiation thus determined is $1.2 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$ (Fig. 1j), close to the value of $(1.5 \pm 0.1) \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$ measured in bulk solution¹⁰. Once initiated, unwinding of the 18-bp duplex proceeds quickly to completion (Fig. 1a, b). The duration of the unwinding events, an average of 0.4 ± 0.2 s, is also consistent with bulk-solution values determined using DNA with and without dye labels^{10,19}. Therefore, neither DNA immobilization nor dye attachment significantly affects DNA unwinding. The characteristics of 18-bp unwinding are independent of [Rep], hence unwinding probably results from the action of a well-defined entity. Several intermediate E_{FRET} values, indicative of partially unwound intermediates, could be resolved during unwinding. Complete unwinding of a 40-bp duplex takes longer, about 1 s if uninterrupted (Fig. 1c, d); however, stalls (defined as interruptions in the E_{FRET} decrease which persist for longer than 1 s) are observed in $\sim 70\%$ of the 40-bp unwinding events (see Fig. 1e–h, and discussion below). Before signal termination, the E_{FRET} for the 40-bp duplex drops to lower values than for the 18-bp duplex, probably because the time-averaged R reaches larger values before complete unwinding.

In order to correlate the unwinding data with the DNA-binding properties of Rep, we designed DNA III (Fig. 2), which should be sensitive to changes in the time-averaged conformation of the

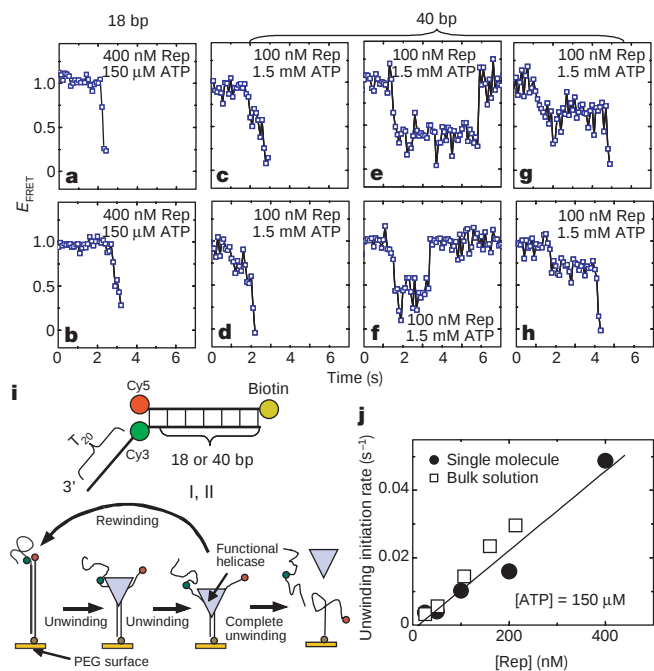


Figure 1 FRET assay for single DNA unwinding. **a–h**, Single DNA unwinding time records (100-ms bins) typical of four distinct patterns: complete unwinding of 18 bp (**a**, **b**), complete unwinding (**c**, **d**), stall and DNA rewinding (**e**, **f**), and stall and unwinding re-initiation (**g**, **h**) of 40 bp. When unwinding is completed (**a–d**, **g–h**), the donor strand quickly diffuses away, abruptly terminating the signal. **i**, The experimental scheme (see text). **j**, The rate of unwinding initiation (the inverse of time between the delivery of enzyme with ATP and the unwinding initiation, averaged over > 50 events for each [Rep]) versus [Rep]. A linear fit and bulk-solution values¹⁰ are shown.

ssDNA tail induced upon Rep binding. In the absence of ATP, we observed two binding conformations of the Rep monomer/DNA complex, with little specificity for the ss/dsDNA junction (Supplementary Information Figs I and II).

In the presence of ATP, if $[Rep] < 2$ nM, DNA unwinding rarely occurs (experiments with DNA I did not detect any unwinding). Hence, under these conditions, we can probe the conformations of DNA III induced by Rep binding. Individual time traces showed brief bursts of fluctuations (Fig. 2a and Supplementary Information Fig. III) that we interpret as reflecting Rep monomer binding, conformational fluctuations, and dissociation. Our interpretation is based on the following observations. (1) The average burst duration (~ 4 s) is independent of $[Rep]$, and matches the monomer dissociation rate (~ 0.24 s $^{-1}$) measured in bulk solution¹⁰. (2) The number of bursts per second is linearly dependent on $[Rep]$ with a slope of 0.7×10^7 M $^{-1}$ s $^{-1}$ (Supplementary Information Fig. III), slightly lower than the bulk solution rate constant ($\sim 2 \times 10^7$ M $^{-1}$ s $^{-1}$)¹⁰ for monomer binding to ssDNA (dT₁₆). This difference is understandable, as intermediate steps appear necessary between the Rep monomer binding and the observation of the burst (see below), and the steps involved in ssDNA translocation may not occur with unit probability. (3) The reciprocal experiment where biotinylated Rep (Rep-BCCP) monomers were immobilized with DNA III (without biotin) and ATP in solution showed similar fluctuations (~ 3 s average duration; Fig. 2b) when DNA binds to the Rep-BCCP monomer, indicating that the fluctuations are not caused by interactions of multiple monomers. In a separate experiment using DNA I (without biotin) and immobilized Rep-BCCP, no unwinding was observed, whereas the DNA binds transiently to the protein for the same average duration.

ATP hydrolysis appears necessary to observe these bursts, as they were not observed when ATP γ S (with almost identical binding kinetics to ATP, but $500 \times$ lower hydrolysis rate; ref. 20) were used. No such fluctuations were observed with AMPPNP, or with an ATPase deficient mutant RepK28I in the presence of ATP²¹. They

appear specific for a 3'-ss/dsDNA junction because a DNA molecule possessing the same 18-bp duplex but with a 5'-(dT)₁₉ tail did not show such bursts. Combined with the evidence that PcrA, a structural homologue of Rep, can translocate as a monomer along ssDNA with a 3' to 5' directional bias at a rate of ~ 50 bases s $^{-1}$ (refs 9, 22), and our bulk-solution observation that Rep monomers appear able to move with the same directionality at a faster rate (C. Fischer, J. Hsieh and T.M.L., unpublished results), we interpret the fluctuations as the end result of Rep monomer's translocation toward the junction. Further evidence that a burst occurs after translocation (as opposed to reflecting translocation itself, which would be too fast to be observed here) is obtained from data on the effect of $[ATP]$ and tail length (Supplementary Information Fig. III).

These fluctuations appear futile—that is, they do not lead to DNA unwinding, as experiments with DNA I under the same conditions did not detect unwinding. Therefore, DNA unwinding is not a simple consequence of a Rep monomer's ability to translocate along ssDNA toward the junction, where it might trap ssDNA formed transiently at the junction via thermal fraying of the duplex. Instead, there is a slow, rate-limiting step required for initiating DNA unwinding after monomer positioning near the junction.

We considered two possibilities for this rate-limiting step. The first is that the Rep monomer must undergo slow conformational transition(s) to a functional monomeric helicase. In this case, once binding near the junction is saturated, the unwinding initiation rate should remain constant even at increased $[Rep]$. The second possibility involves additional protein binding. Then, the unwinding initiation rate would depend on $[Rep]$ even after the junction is saturated; a linear dependence would be expected if two monomers are required, and so on. Saturation of the ss/dsDNA junction by a monomer indeed occurs at $[Rep] \geq 20$ nM (DNA III showed continuous fluctuations similar to the fluctuations within a burst shown in Fig. 2a, probably because a new monomer occupies the junction as soon as the junction becomes unoccupied by the previous one; Supplementary Information Fig. IV). But the unwinding initiation rate continues to increase linearly with $[Rep]$ up to 400 nM (Fig. 1j), favouring the second model. Therefore, the rate-limiting step for DNA unwinding initiation involves the interaction of more than one Rep monomer (our data are consistent with two, but future experiments, using for example dye-labelled proteins, will directly address this question), independently confirming conclusions based on pre-steady-state single-turnover kinetic studies performed in bulk solution¹⁰.

Our single-molecule study further indicates that the Rep monomer indeed uses ATP hydrolysis to position itself near the ss/dsDNA junction, in agreement with the main evidence suggesting that a PcrA monomer has helicase activity^{5,9,22}, yet DNA unwinding is not initiated until the interaction with additional protein(s) activates the unwinding. The present work also provides a more sensitive test for monomer activity because partial unwinding of a small fraction of DNA molecules ($< 3\%$) would be very difficult to detect in a bulk-solution stopped-flow study, whereas even transient unwinding would be detectable from single molecules. In addition, the DNA does not need to be incubated with proteins in the absence of ATP, which is necessary for measuring the early phase of unwinding in bulk solution. We can therefore rule out concerns that a Rep monomer/DNA complex may exist in an inactive form that cannot be rescued by adding ATP afterwards.

The single-molecule studies reported here offer insight into why Rep and many other helicases display only limited unwinding processivities *in vitro*. We observed frequent stalls of the Rep helicase during the course of unwinding a 40-bp duplex (DNA II in Fig. 1). Over two-thirds of the DNA-unwinding events resulted in stalls (Fig. 1e–h) which, in most cases, led to complete DNA rewinding (Fig. 1e, f). The average stall duration was ~ 4 s, longer than the time it takes to unwind the duplex without a stall (~ 1 s, Fig. 1c, d), but

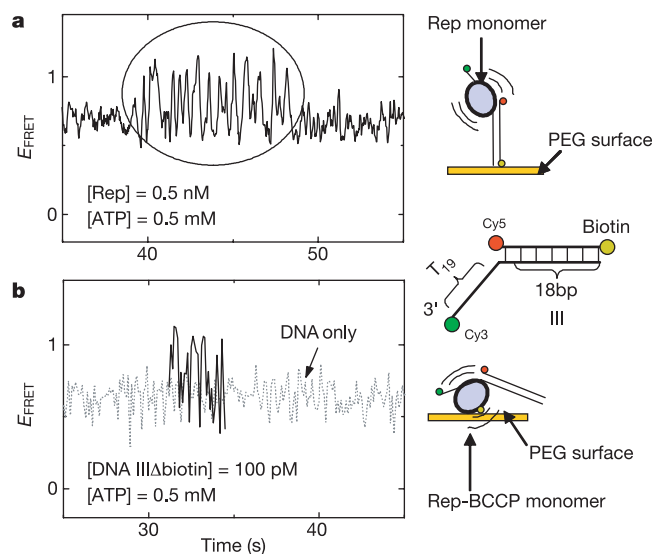


Figure 2 ATP-dependent, junction-specific fluctuations of Rep monomer. **a**, A burst of E_{FRET} fluctuations (circled) is caused by enzyme binding to surface-immobilized DNA III and conformational fluctuations (33-ms bins). **b**, A burst is observed when DNA III (minus biotin) molecules bind to a surface-immobilized, biotinylated Rep (Rep-BCCP) monomer. Fluorescence signal is detectable only during the burst (100-ms bins). Also shown is a typical trace obtained from immobilized DNA III only (grey) for comparison. Accompanying cartoons illustrate Rep monomer conformational fluctuations with either DNA or protein immobilized.

close to the dissociation timescale of a Rep monomer bound at a ss/dsDNA junction (~ 4 s) measured via ATP-dependent, junction-specific fluctuations (Fig. 2a). This suggests that the stalls occur as a result of partial dissociation of a functional helicase, and that the DNA re-anneals only when the remaining bound monomer, which is unable to continue unwinding, dissociates.

This interpretation is further supported by the observation that a fraction of the stalled complexes can re-initiate unwinding, resulting in completion of unwinding (Fig. 1g, h). Re-initiation was not observed in the absence of free Rep protein (that is, Rep in solution, but not associated with any DNA), and the fraction of stalled complexes that re-initiate unwinding increases with increasing concentration of free Rep (Fig. 3a). Hence, the interaction with additional Rep protein(s) appears to be responsible for the restart of stalled complexes. The fraction of stalled complexes that re-initiate unwinding also increases with increasing [ATP] (Fig. 3b). Therefore, an ATP-hydrolysis-driven process, potentially ssDNA translocation of another monomer towards the stall site, may promote functional helicase reassembly.

Overall, our analysis suggests that the limited processivity observed for Rep helicase *in vitro* is due to the relative instability of the functional helicase complex during unwinding. Such instability may contribute to the low unwinding processivities observed for many other helicases *in vitro*. *In vivo*, the relative instability of the helicase complex may enable the regulation of the unwinding activity via other factors that stabilize the complex. For Rep, ϕ X174 gene A protein *in vivo* enables Rep to unwind at least 7,000 bp processively²³. Our observation of stalled monomers also rules out the possibility that a Rep monomer can unwind DNA processively once a functional multi-Rep complex has initiated unwinding.

The cartoon in Fig. 3c summarizes our current view. After binding to a 3' ssDNA tail, a Rep monomer hydrolyses ATP to translocate (3' to 5') towards the ss-dsDNA junction. The mono-

mer displays futile conformational fluctuations either until it dissociates or until additional monomer(s) binds to form a functional helicase that can then initiate unwinding. Unwinding stalls occur if the functional helicase partially dissociates, leaving a monomer at the junction that cannot unwind, thus preventing DNA rewinding; unwinding can resume only if a functional helicase reassembles at the stall site. Future experiments using fluorescently labelled proteins may be able to observe the coupling between the protein conformational changes and unwinding activity by simultaneously measuring them on single DNA molecules. $\square$

Methods

DNA preparation

5'-Cy5-GCCTCGCTGCCGTCGCCA-3' biotin and 5'-TGGCGACGGCAGCGAGGC-Cy3-(T)₂₀-3' for DNA I, 5'-Cy5-GCCCTGCTGCCGACCAACGATGGTTACATTCCCGCTGCTG-3' biotin and 5'-CAGCAGCGGAATGTAACCATCGTTGGTGGGCAGCAGGGC-Cy3-(T)₂₀-3' for DNA II. For DNA III, the second strand of DNA I is replaced by 5'-TGGCGACGGCAGCGAGGC-(T)₁₉-Cy3-T-3'. Cy3 and Cy5 were incorporated in phosphoramidite forms, and biotin was added as BiotinTEG CPG (all three from Glen Research) during oligonucleotide synthesis. DNA was gel purified and annealed in a buffer containing 10 mM Tris, 0.5 M NaCl, pH 8.0.

Protein preparation

Wild-type Rep and Rep-BCCP were purified as described²⁴. Construction of Rep-BCCP is described in Supplementary Information Fig. II legend. The Stopped-flow fluorescence unwinding signal¹⁹ for Rep-BCCP reaches saturation at [Rep-BCCP] = 200 nM and showed an unwinding stepping rate, assuming 4-bp step size, of 11.5 ± 0.5 s⁻¹, in agreement with 14.3 ± 1.1 s⁻¹ for wild-type Rep¹⁰.

Single-molecule measurements

Quartz slides were treated with an amino-silane reagent, Vectabond (Vector), and then incubated with polyethylene glycol (PEG, Shearwater Polymers, containing 20% methoxy-PEG-succinimidyl succinate and 0.25% biotin-PEG-OCH₂CH₂-CO₂-NHS) in 0.1 M sodium bicarbonate (pH 8.3) for 3 h. Streptavidin solution (0.2 mg ml⁻¹) was applied and washed away before addition of biotinylated DNA solution (10–50 pM, Tris 10 mM, NaCl 50 mM, pH 8). Adsorption of Rep protein was reduced by $>1,000\times$ on this surface, compared to bovine serum albumin (BSA)-coated or bare glass, and solution Rep activities were faithfully reproduced. Rep-BCCP was immobilized similarly (details in Supplementary Information Fig. II). Single-molecule measurements were done at room temperature in a buffer containing 20 mM Tris (pH 7.5), 6 mM NaCl, 1.7 mM MgCl₂, 10% glycerol, 0.1 mg ml⁻¹ glucose oxidase, 0.02 mg ml⁻¹ catalase, 1% 2-mercaptoethanol and 3–9% w/w glucose.

Received 20 March; accepted 9 August 2002; doi:10.1038/nature01083.

- Lohman, T. M. & Bjornson, K. P. Mechanisms of helicase-catalyzed DNA unwinding. *Annu. Rev. Biochem.* **65**, 169–214 (1996).
- Hall, M. C. & Matson, S. W. Helicase motifs: the engine that powers DNA unwinding. *Mol. Microbiol.* **34**, 867–877 (1999).
- von Hippel, P. H. & Delagoutte, E. A general model for nucleic acid helicases and their "coupling" within macromolecular machines. *Cell* **104**, 177–190 (2001).
- Patel, S. S. & Picha, K. M. Structure and function of hexameric helicases. *Annu. Rev. Biochem.* **69**, 651–697 (2000).
- Velankar, S. S., Soutanas, P., Dillingham, M. S., Subramanya, H. S. & Wigley, D. B. Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism. *Cell* **97**, 75–84 (1999).
- Mechanic, L. E., Hall, M. C. & Matson, S. W. Escherichia coli DNA helicase II is active as a monomer. *J. Biol. Chem.* **274**, 12488–12498 (1999).
- Levin, M. K. & Patel, S. S. The helicase from hepatitis C virus is active as an oligomer. *J. Biol. Chem.* **274**, 31839–31846 (1999).
- Ali, J. A., Maluf, N. K. & Lohman, T. M. An oligomeric form of E. coli UvrD is required for optimal helicase activity. *J. Mol. Biol.* **293**, 815–834 (1999).
- Dillingham, M. S., Wigley, D. B. & Webb, M. R. Demonstration of unidirectional single-stranded DNA translocation by PcrA helicase: Measurement of step size and translocation speed. *Biochemistry* **39**, 205–212 (2000).
- Cheng, W., Hsieh, J., Brendza, K. M. & Lohman, T. M. E. coli Rep oligomers are required to initiate DNA unwinding *in vitro*. *J. Mol. Biol.* **310**, 327–350 (2001).
- Bianco, P. R. *et al.* Processive translocation and DNA unwinding by individual RecBCD enzyme molecules. *Nature* **409**, 374–378 (2001).
- Dohoney, K. M. & Gelles, J. χ -Sequence recognition and DNA translocation by single RecBCD helicase/nuclease molecules. *Nature* **409**, 370–374 (2001).
- Weiss, S. Measuring conformational dynamics of biomolecules by single molecule fluorescence spectroscopy. *Nature Struct. Biol.* **7**, 724–729 (2000).
- Ha, T. *et al.* Probing the interaction between two single molecules—fluorescence resonance energy transfer between a single donor and a single acceptor. *Proc. Natl Acad. Sci. USA* **93**, 6264–6268 (1996).
- Zhuang, X. W. *et al.* A single-molecule study of RNA catalysis and folding. *Science* **288**, 2048–2051 (2000).
- Ha, T. Single molecule fluorescence resonance energy transfer. *Methods* **25**, 78–86 (2001).
- Sofia, S. J., Premnath, V. & Merrill, E. W. Poly(ethylene oxide) grafted to silicon surfaces: grafting density and protein adsorption. *Macromolecules* **31**, 5059–5070 (1998).

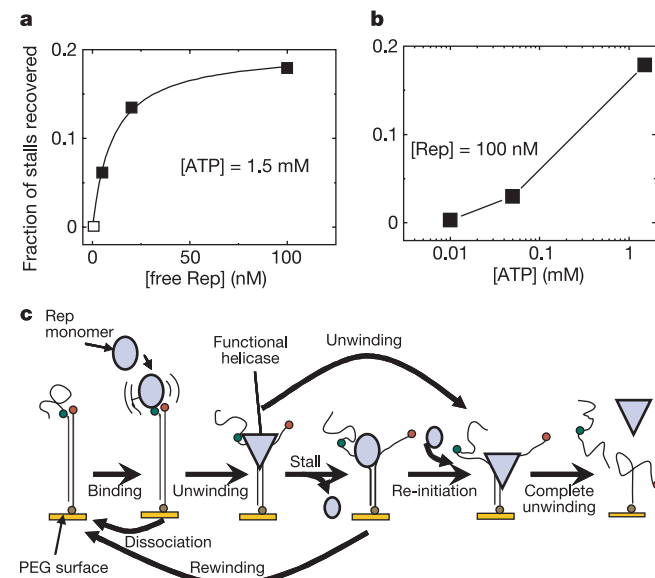


Figure 3 Mechanism of unwinding stall and re-initiation. **a**, Fraction of stalls recovered versus the concentration of free Rep (filled squares). The open square represents the observation that no stall recovery occurred if the unwinding is initiated by adding ATP and trap DNA (5 μ M dT₁₆) onto immobilized DNA II incubated with 200 nM Rep. A fit to the data according to $A/(B + [Rep])$ is also shown ($A = 0.2$, $B = 10.4$ nM). Its saturation may imply the presence of other necessary steps or conformational changes that compete with dissociation. **b**, Fraction of stalls recovered versus [ATP]. At least 100 stalls were analysed for each condition in **a** and **b**. **c**, Unwinding mechanism.

18. Chao, K. & Lohman, T. M. DNA-induced dimerization of the *Escherichia coli* Rep helicase. *J. Mol. Biol.* **221**, 1165–1181 (1991).
19. Bjornson, K. P., Amaratunga, M., Moore, K. J. M. & Lohman, T. M. Single-turnover kinetics of helicase-catalyzed DNA unwinding monitored continuously by fluorescence energy transfer. *Biochemistry* **33**, 14306–14316 (1994).
20. Moore, K. J. M. & Lohman, T. M. Kinetic mechanism of adenine nucleotide binding to and hydrolysis by the *Escherichia coli* Rep monomer. 1. Use of fluorescent nucleotide analogues. *Biochemistry* **33**, 14550–14564 (1994).
21. Wong, I. & Lohman, T. M. A two-site mechanism for ATP hydrolysis by the asymmetric Rep dimer P2S as revealed by site-specific inhibition with ADP-ALF4. *Biochemistry* **36**, 3115–3125 (1997).
22. Dillingham, M. S., Wigley, D. B. & Webb, M. R. Direct measurement of single-stranded DNA translocation by PcrA helicase using the fluorescent base analogue 2-aminopurine. *Biochemistry* **41**, 643–651 (2002).
23. Yarranton, G. T. & Gefter, M. L. Enzyme-catalyzed DNA unwinding: studies on *Escherichia coli* rep protein. *Proc. Natl Acad. Sci. USA* **76**, 1658–1662 (1979).
24. Lohman, T. M., Chao, K., Green, J. M., Sage, S. & Runyon, G. T. Large-scale purification and characterization of the *Escherichia coli* rep gene product. *J. Biol. Chem.* **264**, 10139–10147 (1989).

Supplementary Information accompanies the paper on Nature's website
(<http://www.nature.com/nature>).

Acknowledgements We thank C. Chidsey for suggesting the use of a PEG surface, and T. Ho for synthesis and purification of the oligonucleotides. This work was supported by the NSF and AFOSR (S.C.), by the NIH (T.M.L.), and by a Searle scholars award, the NIH, an NSF CAREER award, the Research Corporation, and the UIUC research board (T.H.)

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.H.
(e-mail: tjha@uiuc.edu).

SATB1 targets chromatin remodelling to regulate genes over long distances

Dag Yasui*, Masaru Miyano*, Shutao Cai*, Patrick Varga-Weisz† & Terumi Kohwi-Shigematsu*

* Life Sciences Division, Lawrence Berkeley National Laboratory, University of California, Berkeley 94720, USA

† Marie Curie Research Institute, The Chart, Oxted, Surrey RH8 OTL, UK

Eukaryotic chromosomes are organized inside the nucleus in such a way that only a subset of the genome is expressed in any given cell type, but the details of this organization are largely unknown^{1–3}. SATB1 ('special AT-rich sequence binding 1'), a protein found predominantly in thymocytes⁴, regulates genes by folding chromatin into loop domains, tethering specialized DNA elements to an SATB1 network structure⁵. Ablation of SATB1 by gene targeting results in temporal and spatial mis-expression of numerous genes and arrested T-cell development, suggesting that SATB1 is a cell-type specific global gene regulator⁶. Here we show that SATB1 targets chromatin remodelling to the *IL-2Rα* ('interleukin-2 receptor α') gene, which is ectopically transcribed in SATB1 null thymocytes. SATB1 recruits the histone deacetylase contained in the NURD chromatin remodelling complex to a SATB1-bound site in the *IL-2Rα* locus, and mediates the specific deacetylation of histones in a large domain within the locus. SATB1 also targets ACF1 and ISWI, subunits of CHRAC and ACF nucleosome mobilizing complexes, to this specific site and regulates nucleosome positioning over seven kilobases. SATB1 defines a class of transcriptional regulators that function as a 'landing platform' for several chromatin remodelling enzymes and hence regulate large chromatin domains.

SATB1 forms a three-dimensional network structure in mouse thymocyte nuclei⁵. It specifically binds to double-stranded DNA with a specialized ATC sequence context that readily becomes base-

unpaired in supercoiled DNA^{4,7,8}. Such base unpairing regions (BURs) are important elements of matrix attachment regions^{7,8}. SATB1 regulates genes located as much as ~50 kilobases (50 kb) away from the attached sites. In SATB1 null thymocytes, which arrest at the CD4⁺CD8⁺ double positive stage, the normal SATB1-bound genomic loci are detached and associated genes are dysregulated⁵. What is the mechanism that connects SATB1 function in nuclear organization to transcription? We tested a hypothesis that the SATB1 network provides assembly sites for chromatin remodelling and modifying factors involved in gene regulation and other nuclear functions⁹. ATP-dependent chromatin remodelling factors use the energy gained by ATP hydrolysis to alter nucleosome array structures^{2,10,11}. Histone modifying enzymes are involved in both transcriptional activation (such as histone acetylation) and repression (for example, histone deacetylation)⁹.

Using DNA affinity chromatography designed to purify SATB1, we investigated whether chromatin remodelling and modifying factors specifically co-purify with SATB1. We compared elution profiles of extracts prepared from wild-type and SATB1 null thymocytes (+/+ and -/-, respectively). A BUR containing a core unwinding element found in a matrix attachment region 3' of the immunoglobulin heavy chain (IgH) enhancer was multimerized and used to make a DNA affinity column to purify SATB1. In order to trap non-specific A + T-rich and other DNA binding proteins, thymocyte extracts were first loaded onto a mutated (Mut) BUR column that has lost the unwinding propensity, but retains the A + T-rich feature. The flow-through of this column was applied onto the wild-type (WT) BUR column. As shown in Fig. 1a (middle panel, lanes, SM, Mut, WT), SATB1 flows through the mutant BUR affinity column and is almost completely retained in the wild-type BUR column¹². SATB1 was eluted from this column by 0.4–0.8 M [Na⁺]. We asked whether the NURD (nucleosome remodelling and histone deacetylase) complex co-purifies with SATB1. The NURD complex has been implicated in transcriptional repression of several genes¹³, and contains ATP-dependent remodelling enzyme Mi-2 (relative molecular mass 240,000; *M_r* 240K), histone deacetylase 1 (HDAC1; 62K) and 2 (HDAC2; 55K), histone deacetylase associated co-repressor mSin3A (160K) and MTA-2 (70K)^{14–16}. In addition, we monitored the presence of nucleosome-dependent ISWI-ATPase (140K), and ACF1 (180K), both subunits of the ACF and CHRAC complexes (reviewed in ref. 17). These complexes are known to mobilize nucleosomes to reorganize nucleosomal arrays over large distances *in vitro*¹⁷. However, nothing is known about their *in vivo* function, and whether they are targeted to specific sites. In the fractions that contained SATB1, the NURD complex subunits Mi-2, mSin3A, MTA-2, HDAC1 and HDAC2 and CHRAC/ACF complex subunits ACF1 and ISWI, were also detected (Fig. 1a, top panel). These chromatin remodelling factors were retained on the BUR column via SATB1, as none of these proteins were retained when SATB1 deficient extracts were applied (Fig. 1b, top panel). Mi-2, ACF1, mSin3A, ISWI, MTA-2, HDAC1 and HDAC2 proteins were present at roughly the same abundance in a SATB1 deficient thymic extract as the wild-type extract (Fig. 1a, b, SM, Mut and WT lanes). Apparently, chromatin remodelling factors that co-purified with SATB1 represent a subfraction of each of these factors present in thymocytes, as these proteins were also found in the flow-through fraction of the BUR affinity column (Fig. 1a, lane WT). Poly (ADP-ribose) polymerase (PARP), another BUR-binding protein¹⁸ found in both wild-type and SATB1 null thymocyte extracts, was retained in the BUR affinity column (Fig. 1a, b, bottom panels), indicating that the affinity column used for both extracts was active. The data also indicate that unlike SATB1, a BUR-binding protein such as PARP is not sufficient to recruit chromatin remodelling factors to sites of the specialized DNA context. The elution profile of chromatin remodelling factors did not exactly match that of SATB1. One explanation could be that these factors associate with SATB1 less strongly than SATB1 to DNA. Co-elution of chromatin

remodelling factors along with SATB1 is not an artifact of SATB1 potentially being a 'sticky' protein, because other factors such as retinoblastoma protein (Rb) were absent in the eluted fractions (data not shown).

Do chromatin remodelling factors form protein complexes with SATB1 in the absence of DNA? Anti-SATB1 antibody co-immunoprecipitated ISWI, Mi-2, MTA-2 and HDAC1 along with SATB1 from wild-type but not SATB1 null thymic extracts that lack any genomic DNA (Fig. 2a, lanes 1 and 2). PARP, Rb and NF κ B proteins were not detected in the anti-SATB1 immunoprecipitate from the wild-type extracts (Fig. 2a, lane 1 and data not shown). Control antibody (monoclonal anti-p27 antibody) fails to immunoprecipitate any of these proteins (Fig. 2a, lane 3). In a separate experiment, anti-HDAC1 antibody immunoprecipitated SATB1 only from wild-type extract, in addition to mSin3A, MTA-2 and HDAC1 itself (Fig. 2b, lanes 1 and 2). These data show that SATB1 assembles protein complexes containing factors found in CHRAC/ACF and NURD complexes, including ATP-dependent remodelling enzymes, histone deacetylases and their associating factors in the absence of DNA.

We examined whether SATB1 directly interacts with the components of CHRAC/ACF using purified recombinant ISWI isoform human SNF2H, human ACF1, the CHRAC-specific histone-fold proteins p15 and p17 (ref. 19), and SATB1 (amino acids 90–763). By immunoprecipitation with antibodies against each protein, we show that SATB1, hACF1 and hSNF2H make a protein complex *in vitro* (Fig. 2c, lanes 1–4). Histone-fold proteins were dispensable for the formation of this complex. SATB1 directly interacted separately with both hSNF2H and hACF1 (Fig. 2c, lanes 5–7 and 8–10). Truncated SATB1 (amino acids 365–763) failed to interact with either hSNF2H or hACF1 (Fig. 2c, lanes 11–14), suggesting that the SATB1 domain of amino acids 90–365 contain an hSNF2H binding site(s). Consistent with the data from BUR DNA affinity chromatography, PARP did not interact with CHRAC/ACF (Fig. 2c, lanes 15–18). SATB1 is a targeting factor of CHRAC/ACF.

We next examined whether SATB1 directs chromatin remodelling complexes to the SATB1 binding loci *in vivo*. We first studied HDAC1, a component of the NURD complex. This was done by chromatin immunoprecipitation (ChIP) assays. Briefly, thymocyte chromatin crosslinked with formaldehyde, purified with a urea

gradient, and digested with restriction enzymes, was immunoprecipitated with either anti-SATB1 or anti-HDAC1 antibody^{20,21}. We focused on the *IL-2R α* locus because the *IL-2R α* gene is ectopically transcribed in SATB1 null double positive thymocytes⁶. The double positive subset represents >84% of the total thymocyte population. Within the *IL-2R α* locus, we identified a 700-bp fragment that exhibits very high affinity ($K_d = 0.1$ – 0.5 nM, data not shown), to SATB1 *in vitro* (SATB1-bound sequence 700; SBS700) which mapped to the first intron, 6.8–7.5 kb downstream of the transcriptional start site (Fig. 3c). There is no additional SATB1-bound site

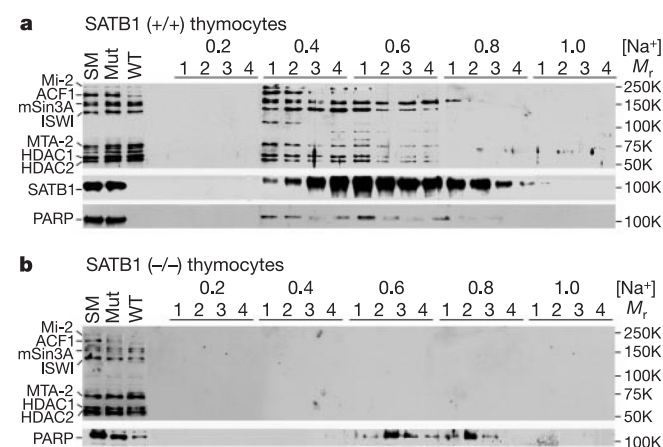


Figure 1 Chromatin remodelling factors co-purify with SATB1 in DNA affinity chromatography. **a**, Proteins in SATB1 (+/+) thymic extract (SM), mutant BUR (base unpairing region) column flow-through (Mut), wild-type BUR column flow-through (WT) and proteins eluted with increasing concentrations of NaCl were identified by western analysis. Mi-2, ACF1, mSin3A, ISWI, MTA-2, HDAC1 and HDAC2 signals are shown in the top panel. SATB1 signal is shown in the middle panel and PARP, another BUR binding protein, is shown in the lower panel. **b**, When BUR affinity chromatography was performed with SATB1 (-/-) extract, only PARP was eluted (lower panel).

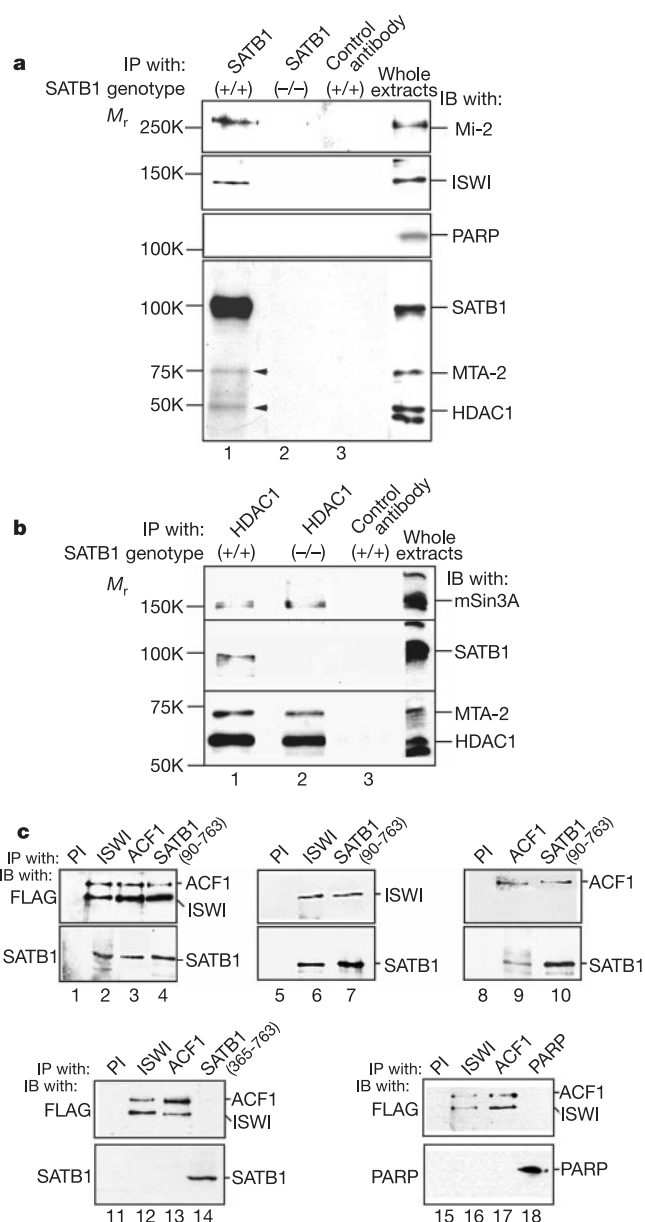


Figure 2 SATB1 directly interacts with several chromatin remodelling factors. Proteins immunoprecipitated (IP) with **a**, anti-SATB1 antibody or **b**, anti-HDAC1 from SATB1 (+/+) (lane 1) and SATB1 (-/-) thymic extracts (lane 2) were identified by western blot with antibodies as indicated. In both **a** and **b**, lane 3 shows control antibody (anti-p27 antibody) immunoprecipitation of SATB1 (+/+) extract, and lane 4 shows proteins in starting material. **c**, Protein immunoprecipitated with anti-ISWI, anti-ACF1, anti-SATB1, and anti-PARP antibodies from mixtures of purified FLAG-fused recombinant ISWI (hSNF2H), ACF1 (hACF1), his-tag-fused SATB1 (amino acids 90–763), truncated SATB1 (amino acids 365–763), and full-length PARP were identified by western blot with anti-FLAG, anti-SATB1 and anti-PARP antibodies. PI, preimmune serum; IP, immunoprecipitation; IB, immunoblot.

identified within ~16 kb upstream of SBS700. We found that SBS700 is indeed bound to SATB1 as well as to HDAC1 *in vivo* in wild-type thymocytes (Fig. 3a, upper row, anti-SATB1 and anti-HDAC1; lane 3). In contrast, SBS700 was not bound to HDAC1 in SATB1 null thymocytes (Fig. 3a, lower row, anti-HDAC1; lane 3). This was revealed by polymerase chain reaction (PCR) amplification with a specific primer set designed for SBS700 and regardless of the precise number of cycles used in the range of 25 to 30 cycles. An intronic sequence 2.8 kb downstream of exon 1, which was not bound to SATB1 in wild-type thymocytes (Fig. 3a, upper row, anti-SATB1, lane 2) was also not bound to HDAC1 in both wild-type and null thymocytes (Fig. 3a, upper and lower rows, anti-HDAC1; lane 2). As a positive control, we used SBS336, which was cloned as a SATB1-bound sequence⁵ from total *in vivo* crosslinked genomic DNA (Fig. 3a, upper and lower rows, anti-SATB1; lane 4). Like SBS700, SBS336 was also found in anti-HDAC1 antibody immunoprecipitates from wild-type but not SATB1 null thymocytes (Fig. 3a, upper and lower rows, anti-HDAC1, lane 4). Using a specific primer set that amplifies the *IL-2R α* promoter region, a faint PCR product was seen in wild-type thymocytes (Fig. 3a upper row, anti-HDAC1; lane 1), even though the first 3-kb promoter region totally lack SATB1 binding sequences. Such a band was absent in null thymocytes (Fig. 3a, lower row, anti-HDAC1; lane 1). The non-immune serum precipitated DNA pool never gave rise to a PCR amplified signal under the same conditions (Fig. 3a, upper and lower rows, non-immune serum; lanes 1–4). These data show that SATB1 is required for recruitment of HDAC1 onto the SATB1's target sequence and the promoter within the *IL-2R α* locus as well as other loci.

Acetylation of histones in nucleosomes indicates transcriptionally active chromatin^{22,23}. HDAC1 specifically recruited to the SATB1-binding site in the *IL-2R α* intron in thymocytes is expected to have a role in controlling the histone acetylation status of that locus. Therefore, we immunoprecipitated *in vivo* crosslinked chromatin fragments with anti-acetylated histone H4 antibody. SBS700 is absent in this immunoprecipitated chromatin from wild-type thymocytes. In contrast, in SATB1 null thymocytes where SBS700 is

not bound to HDAC1 *in vivo*, histones in the SBS700 region are acetylated (Fig. 3a, upper and lower rows; anti-H4ace; lane 3), consistent with a transcriptionally active state. Furthermore, in SATB1 null thymocytes, histones in the SBS700 distal promoter region (Fig. 3a, lower row, anti-H4 ace; lane 1) and the remote intronic sequence (Fig. 3a, lower row, anti-H4 ace; lane 2) are also hyperacetylated, which is consistent with the ectopic transcription of *IL-2R α* in the absence of SATB1. Identical results on histone acetylation were obtained with the SBS336 locus. Therefore, SATB1 affects chromatin structure over distance in the *IL-2R α* gene locus by recruiting HDAC1 to specific DNA sequences, thus maintaining a hypoacetylated nucleosome status, and packaging chromatin into a transcriptionally inactive state.

Similar to HDAC1, we also found by chromatin immunoprecipitation assay that both ISWI and ACF1 were specifically bound *in vivo* to SBS700 of the *IL-2R α* locus as well as to SBS336 in wild-type thymocytes but not in SATB1 null thymocytes (Fig. 3b, upper and lower rows, anti-ACF1 and anti-ISWI; lanes 3 and 4). Furthermore, both factors were also found associated with the *IL-2R α* promoter only when SATB1 is present (Fig. 3b, anti-ACF1 and anti-ISWI; lane 1), but never to the intronic sequence regardless of the presence of SATB1 (Fig. 3b, upper row, anti-ACF1 and anti-ISWI; lane 2). In budding yeast, a related ISWI-complex acts in transcriptional repression by creating an inaccessible local chromatin structure at several promoters (reviewed in ref. 17). Our data show that SATB1 is necessary for both the proper repression of the *IL-2R α* gene in double positive thymocytes⁶ and the recruitment of ISWI and ACF1 to a specific site within the gene locus, implicating ISWI and ACF1 in repression of *IL-2R α* .

The biological consequence of recruiting ATP-dependent chromatin remodelling enzymes such as ISWI to the *IL-2R α* locus may be reflected by changes in nucleosome positioning in this locus. Therefore, we performed micrococcal nuclease (MNase) digestion in isolated thymocyte nuclei followed by Southern hybridization, and compared results between wild-type and SATB1 null cells. In each region examined, we also performed micrococcal nuclease

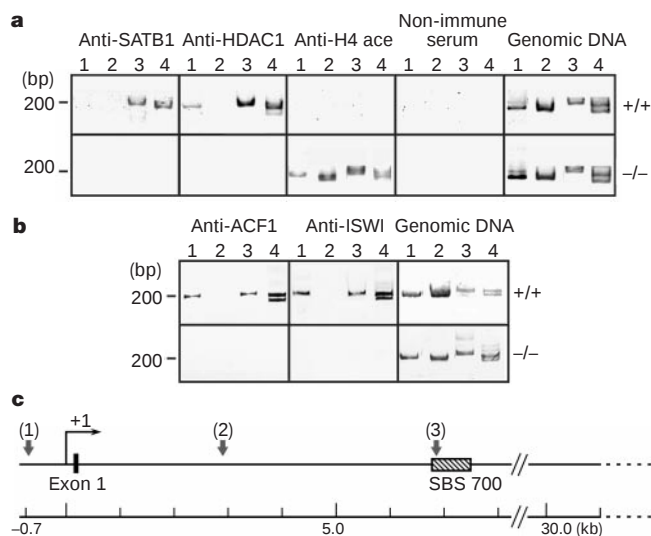


Figure 3 SATB1 targets HDAC1, ISWI and ACF1 to the *IL-2R α* locus *in vivo*. PCR amplification of DNA isolated from immunoprecipitated SATB1 wild-type (+/+) chromatin (upper row) and SATB1 (-/-) chromatin (lower row) using antibodies to **a**, SATB1, HDAC1, acetylated histone H4 (H4ace), control non-immune serum and **b**, to ACF1, ISWI, or genomic DNA. Each DNA sample was amplified with four primer pairs; *IL-2R α* promoter (lanes 1), *IL-2R α* intron (lanes 2), *IL-2R α* SBS700 (lanes 3) or SBS336 (lanes 4). DNA size markers are shown to the left of the panel. **c**, A map of *IL-2R α* with positions of the primers used for (1) promoter, (2) intron, and (3) SBS700 sequences.

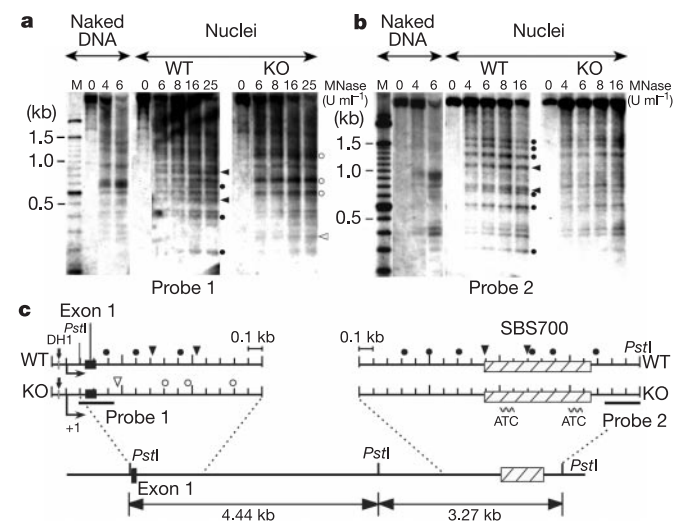


Figure 4 Alteration of nucleosomal positioning in the *IL-2R α* locus in SATB1 null thymocytes. **a**, MNase cleavage sites downstream of the first exon region in wild type (WT), SATB1 null (KO) thymocyte nuclei and naked DNA determined by Southern hybridization of *Pst*I-digested DNA with probe 1. **b**, MNase cleavage sites at the intronic SBS700 determined by Southern hybridization of *Pst*I-digested DNA with probe 2. **c**, Summary map of the *IL-2R α* gene locus. Location of SBS700 region (hatched boxes) and ATC sequences that binds SATB1 (wavy lines) are indicated. For **c**, MNase cleavage sites restricted to WT (solid arrowheads) or KO nuclei (open arrowheads), and enhanced cleavages for WT (solid circle) or KO nuclei (open circles) are shown.

digestion directly on purified genomic DNA to distinguish the cleavage sites of this enzyme at its preferred DNA sequences from its cleavages at nucleosomal linker sites. In a region overlapping the 5' half of SBS700, which contains one of the two SATB1-binding sequences (~100 bp each), we found a clear difference in the pattern of cleavages by micrococcal nuclease between wild-type and SATB1 null nuclei, reflecting changes in nucleosomal distribution (Fig. 4b, arrowheads). We also examined nucleosome positioning in the *IL-2R α* promoter/first exon region which is ~8 kb upstream of SBS700. The promoter region is known to contain two major nuclease hypersensitive regions (DH1 and DH2)²⁴. Although DH1, which is proximal to the promoter, is retained even in the absence of SATB1 (data not shown), nucleosome positioning toward exon 1 also exhibits changes. A major alteration in nucleosome positioning between wild-type and SATB1 null nuclei was revealed in intronic sequences immediately downstream of exon 1, located 0.2–1 kb downstream of the *Pst*I site using probe 1 (Fig. 4a). Because this region, as well as the remaining 7 kb intronic region upstream of SBS700, lacks any ATC sequence cluster and fails to bind SATB1 *in vitro*, any change in nucleosome positioning close to exon 1 is not due to the potential alteration in chromatin structure caused by SATB1 binding. These results show that changes in nucleosomal positioning due to SATB1 ablation can occur as far as 7 kb away from the actual SATB1-binding sequences (Fig. 4c).

The results shown here provide mechanistic information for previously reported activities of matrix attachment regions surrounding the intronic immunoglobulin μ enhancer, which confer enhanced chromatin accessibility to distal positions and modulate chromatin structure^{25–29}. SATB1 links higher-order packaging of chromatin to gene regulation by targeting chromatin remodelling to the bases of chromatin loop domains, where DNA is anchored to the SATB1 network, thereby regulating chromatin structure and gene expression over long distances. We propose that the action of SATB1 represents a general guiding process by which multiple chromatin remodelling factors find entry sites into chromatin at regions with high base-unpairing propensity, providing a mechanism for global gene regulation in higher eukaryotes. □

Methods

BUR affinity chromatography

Ligated, double-stranded oligonucleotides bearing either wild-type (upper strand, 5'-TCTTTAATTTCTAATATATTAGAAATc-3') or mutant DNA (upper strand, 5'-TCTTTAATTTCTACTGCTTTAGAAATc-3') were coupled to Sepharose CL-4B beads (Amersham Pharmacia). Proteins were extracted from 5×10^8 thymocytes in 0.4 M NaCl buffer containing Complete Mini (Roche) protease inhibitor tablets. Extracts were diluted to 0.1 M NaCl, incubated with $10 \mu\text{g ml}^{-1}$ salmon sperm DNA, and passed first through the mutant BUR column then twice through the wild-type BUR column. The wild-type BUR column was washed with 0.1 M NaCl buffer before elution with buffers of increasing NaCl concentration²¹.

Immunoprecipitation

50 μg of protein in Dignam extract (100 μl) was diluted 1 to 1 with 0.2% NP-40/PBS, pre-cleared with rabbit IgG (Santa Cruz) and protein A/G beads (Pierce), then incubated with either anti-SATB1 monoclonal antibody (1:25, Transduction Labs), anti-HDAC1 (1:20, Novus) or an equivalent amount of control antibody (1:25, anti-p27 antibody, Santa Cruz). For *in vitro* co-immunoprecipitation, recombinant FLAG-tagged human SNF2H and hACF1, expressed and purified as described³⁰, and recombinant human CHRAC15/17¹⁹ was used. His-tagged SATB1 (90–763), SATB1 (365–763) and full-length PARP were expressed and purified using standard protocols (Qiagen). 15 ng of each protein was added to each reaction and incubated for 30 min at 4 °C in IP-150 buffer (20 mM HEPES, 150 mM KCl, 1 mM MgCl₂, 0.05% NP-40, 10% glycerol, 0.2 mg ovalbumin, 25 U benzonase), with either anti-ACF1 (1:100), anti-ISWI (1:100)¹⁹, rabbit anti-SATB1^{4,7,8} (1:100), anti-PARP (1:100, Santa Cruz) or pre-immune serum. The rabbit polyclonal antibody against mouse ACF1 was raised against the peptide SPSTVDQVSTPLAAKSR1 at the carboxy terminus of mACF1. Protein-antibody complexes were recovered with protein A/G beads, and washed in PBS containing 0.1% NP-40. Soluble proteins were obtained in Laemmli sample buffer at 95 °C, 5 min, and identified by western analysis with anti-FLAG (M2, Sigma), anti-MTA-2 and anti-Mi-2 (provided by Y. Zhang) and antibodies used in immunoprecipitation.

Chromatin immunoprecipitation

Crosslinked chromatin was prepared from thymocytes and purified by urea gradient centrifugation as described^{4,20}. For each experiment, 150–200 μg of chromatin was

digested with *Sau*3AI (New England Biolab) adjusted to 1.0% NP-40 and pre-cleared first by incubation with protein A/G beads alone, then non-immune rabbit serum and protein A/G beads. Pre-cleared chromatin was divided (~30 μg per tube) and incubated overnight with 5 μl of either rabbit anti-SATB1, anti-HDAC1 (Santa Cruz), anti-acetylated histone H4 (Upstate Biotech), anti-ACF1, affinity-purified anti-ISWI or pre-immune serum. Complexes on beads were washed, then digested with 250 $\mu\text{g ml}^{-1}$ proteinase K and subjected to phenol/chloroform extraction before ethanol precipitation with glycogen. One-twentieth of the DNA from each pool was PCR amplified in reactions containing 1.0 unit of AmpliTaq Gold (Roche), 10 mM Tris-HCl, 1.5 mM MgCl₂ and 1 μM of either: SBS700 matrix attachment region 5'-GCTTACTGGTCTACAGCAAAAT-3' and 5'-CATGCAAAAGTCTTATGCCTGAGC-3', *IL-2R α* intron 5'-TACTGCCACACTAGAAGTTCCG-3' and 5'-CAGGCATAAGTTCCATATCAGGC-3', SBS336 5'-TCCTTACACA CAAAGTGGGCTG-3' and 5'-GGAGTCTTAGATGAGTTGGCATTGC-3' or *IL-2R α* promoter 5'-GACACTATGAGAGAAGGCAAGGG-3' and 5'-ACACCTCGGTA TTGGTTCACTCC-3' using 1 cycle of 95 °C for 10 min, 25–30 cycles of 95 °C for 45 s, 60 °C for 60 s and 72 °C for 90 s. PCR products were resolved by acrylamide gel electrophoresis, stained with Sybr Gold (Molecular Probes) and imaged on a Storm scanner (Molecular Dynamics).

MNase digestion of nuclei and Southern hybridization

Thymic nuclei were prepared in 3.75 mM Tris-HCl, pH 7.6, 0.5 mM EDTA, 20 mM KCl, 1 mM NaHCO₃, 0.05 mM spermine, 0.125 mM spermidine, 0.5 mM dithiothreitol (DTT), 1 tablet of protease inhibitor cocktail per 10 ml, 0.1% NP-40. Nuclei were suspended to 4.5×10^7 nuclei ml⁻¹ in 20 mM HEPES, pH 7.6, 20 mM KCl, 0.5 mM EDTA, 1 mM NaHCO₃, 3 mM CaCl₂. 4.5×10^6 nuclei in the 100- μl buffer were digested with MNase (Roche) at 4, 8 and 16 and 25 U ml⁻¹ concentration for 4 min at 20 °C. Nuclease digestion was stopped by the addition of 1/10 volume of (0.5% SDS, 60 mM EDTA, 3 mg ml⁻¹ proteinase K). MNase-treated genomic DNA was purified and digested with *Bgl*II or *Pst*I and separated on a 1.3% agarose gel, alkali blotted onto Hybond-N⁺ membrane (Amersham) and ultraviolet cross-linked (Stratalinker; Stratagene). To prepare DNA fragments for hybridization probes, PCR reactions was performed with genomic DNA with the following sets of primers: For probe 1: 5'-GAGAATTTTCATCCAGTTCCC-3' and 5'-AGAACCCATGATTCAACTCC-3', for probe 2: 5'-CTCTCCAGTTCAGAGAAAC-3' and 5'-CTGCAGTGTAGGTTGAGAAC-3'. Blotted DNA was hybridized at 43 °C overnight with probes labelled with $\alpha^{32}\text{P}$ -dCTP by Klenow large fragment.

Received 5 June; accepted 13 August 2002; doi:10.1038/nature01084.

- Felsenfeld, G. Chromatin unfolds. *Cell* **86**, 13–19 (1996).
- Travers, A. An engine for nucleosome remodeling. *Cell* **96**, 311–314 (1999).
- Bulger, M. & Groudine, M. Looping versus linking: toward a model for long-distance gene activation. *Genes Dev.* **13**, 2465–2477 (1999).
- Dickinson, L. A., Joh, T., Kohwi, Y. & Kohwi-Shigematsu, T. A tissue-specific MAR/SAR DNA-binding protein with unusual binding site recognition. *Cell* **70**, 631–645 (1992).
- Cai, S., Han, H.-J. & Kohwi-Shigematsu, T. SATB1 creates chromatin domains with specific histone modifications to regulate genes. *Cell* (submitted).
- Alvarez, J. D. et al. The MAR-binding protein SATB1 orchestrates temporal and spatial expression of multiple genes during T-cell development. *Genes Dev.* **14**, 521–535 (2000).
- Kohwi-Shigematsu, T. & Kohwi, Y. Torsional stress stabilizes extended base unpairing in suppressor sites flanking immunoglobulin heavy chain enhancer. *Biochemistry* **29**, 9551–9560 (1990).
- Bode, J. et al. Biological significance of unwinding capability of nuclear matrix-associating DNAs. *Science* **255**, 195–197 (1992).
- Kingston, R. E. & Narlikar, G. J. ATP-dependent remodeling and acetylation as regulators of chromatin fluidity. *Genes Dev.* **13**, 2339–2352 (1999).
- Varga-Weisz, P. ATP-dependent chromatin remodeling factors: nucleosome shufflers with many missions. *Oncogene* **20**, 3076–3085 (2001).
- Fry, C. J. & Peterson, C. L. Chromatin remodeling enzymes: who's on first? *Curr. Biol.* **11**, R185–R197 (2001).
- Dickinson, L. A. & Kohwi-Shigematsu, T. Nucleolin is a matrix attachment region DNA-binding protein that specifically recognizes a region with high base-unpairing potential. *Mol. Cell Biol.* **15**, 456–465 (1995).
- Knoepfler, P. S. & Eisenman, R. N. Sin meets NuRD and other tails of repression. *Cell* **99**, 447–450 (1999).
- Zhang, Y., LeRoy, G., Seelig, H. P., Lane, W. S. & Reinberg, D. The dermatomyositis-specific autoantigen M2 is a component of a complex containing histone deacetylase and nucleosome remodeling activities. *Cell* **95**, 279–289 (1998).
- Taunton, J., Hassig, C. A. & Schreiber, S. L. A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. *Science* **272**, 408–411 (1996).
- Xue, Y. et al. NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. *Mol. Cell* **2**, 851–861 (1998).
- Langst, G. & Becker, P. B. Nucleosome mobilization and positioning by ISWI-containing chromatin-remodeling factors. *J. Cell Sci.* **114**, 2561–2568 (2001).
- Galande, S. & Kohwi-Shigematsu, T. Poly(ADP-ribose) polymerase and Ku autoantigen form a complex and synergistically bind to matrix attachment sequences. *J. Biol. Chem.* **274**, 20521–20528 (1999).
- Poot, R. A. et al. HuCHRAC, a human ISWI chromatin remodelling complex contains hACF1 and two novel histone-fold proteins. *EMBO J.* **19**, 3377–3387 (2000).
- de Belle, I., Cai, S. & Kohwi-Shigematsu, T. The genomic sequences bound to special AT-rich sequence-binding protein 1 (SATB1) *in vivo* in Jurkat T cells are tightly associated with the nuclear matrix at the bases of the chromatin loops. *J. Cell Biol.* **141**, 335–348 (1998).
- Kohwi-Shigematsu, T., deBelle, I., Dickinson, L. A., Galand, S. & Kohwi, Y. Identification of base-unpairing region-binding proteins and characterization of their *in vivo* binding sequences. *Methods Cell Biol.* **53**, 323–354 (1998).
- Kouzarides, T. Histone acetylases and deacetylases in cell proliferation. *Curr. Opin. Genet. Dev.* **9**, 40–48 (1999).

23. Roth, S. Y., Denu, J. M. & Allis, C. D. Histone acetyltransferases. *Annu. Rev. Biochem.* **70**, 81–120 (2001).
24. Soldaini, E. *et al.* Mouse interleukin-2 receptor alpha gene expression. Delimitation of cis-acting regulatory elements in transgenic mice and by mapping of DNase-I hypersensitive sites. *J. Biol. Chem.* **270**, 10733–10742 (1995).
25. Fernandez, L. A., Winkler, M. & Grosschedl, R. Matrix attachment region-dependent function of the immunoglobulin μ enhancer involves histone acetylation at a distance without changes in enhancer occupancy. *Mol. Cell Biol.* **21**, 196–208 (2001).
26. Kirillov, A. *et al.* A role for nuclear NF- κ B in B-cell-specific demethylation of the Igk locus. *Nature Genet.* **13**, 435–441 (1996).
27. Lichtenstein, M., Keini, G., Cedar, H. & Bergman, Y. B cell-specific demethylation: a novel role for the intronic κ chain enhancer sequence. *Cell* **76**, 913–923 (1994).
28. Jenuwein, T. *et al.* Extension of chromatin accessibility by nuclear matrix attachment regions. *Nature* **385**, 269–272 (1997).
29. Forrester, W. C., Fernandez, L. A. & Grosschedl, R. Nuclear matrix attachment regions antagonize methylation-dependent repression of long-range enhancer-promoter interactions. *Genes Dev.* **13**, 3003–3014 (1999).
30. Collins, N. *et al.* An ACF1-ISWI nucleosome remodelling complex is required for DNA replication through condensed chromatin in mammalian cells. *Nature Genet.* (in the press).

Acknowledgements We thank the Joint Genome Institute, especially K. Kadner, for sequencing the BAC clone containing the *IL-2R α* locus; Y. Zhang for antibodies against Mi-2 and MTA-2; H.-J. Han and S. Galande for DNA constructs; M. Grimaldi and I. Kukimoto for expressing recombinant SNF2H and hACF1; and C. García-Jiménez for affinity purification of antibodies. We also thank C. Goding and M. Kohwi for comments on the manuscript. This work was supported by the National Institutes of Health (T.K.-S.), the Marie Curie Cancer Care foundation (P.V.-W.) and training grant AG00266 from the National Institute on Aging (D.Y.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.K.-S. (e-mail: terumiks@lbl.gov).

Reaction path of protein farnesyltransferase at atomic resolution

Stephen B. Long, Patrick J. Casey & Lorena S. Beese

Department of Biochemistry, Duke University Medical Center, Durham, North Carolina 27710, USA

Protein farnesyltransferase (FTase) catalyses the attachment of a farnesyl lipid group to numerous essential signal transduction proteins, including members of the Ras superfamily¹. The farnesylation of Ras oncoproteins, which are associated with 30% of human cancers, is essential for their transforming activity². FTase inhibitors are currently in clinical trials for the treatment of cancer^{2–4}. Here we present a complete series of structures representing the major steps along the reaction coordinate of this enzyme. From these observations can be deduced the determinants of substrate specificity and an unusual mechanism in which product release requires binding of substrate, analogous to classically processive enzymes. A structural model for the transition state consistent with previous mechanistic studies was also constructed. The processive nature of the reaction suggests the structural basis for the successive addition of two prenyl groups to Rab proteins by the homologous enzyme geranylgeranyltransferase type-II. Finally, known FTase inhibitors seem to differ in their mechanism of inhibiting the enzyme.

Protein farnesyltransferase (FTase) and geranylgeranyltransferase types I and II (GGTase-I and GGTase-II) constitute the protein prenyltransferase family of lipid-modifying enzymes¹. These enzymes catalyse the formation of thioether linkages between the C₁ atom of farnesyl (15-carbon by FTase) or geranylgeranyl (20-carbon by GGTase-I, II) isoprenoid lipids and cysteine residues at or near the carboxy-terminus of protein acceptors. Protein substrates

of the prenyltransferases include: Ras, Rho, Rab, other Ras-related small GTP-binding proteins, γ -subunits of heterotrimeric G-proteins, nuclear lamins, centromeric proteins, and many proteins involved in visual signal transduction^{1–3}. The attached lipid is required for proper functioning of the modified protein by mediating membrane associations and specific protein–protein interactions. Inhibitors of FTase cause tumour regression in animals and are currently being evaluated in clinical trials for the treatment of human cancer^{2–4}. FTase and GGTase-I, which are collectively known as the CaaX prenyltransferases, attach their respective isoprenoid to the cysteine residue of a C-terminal CaaX motif (C, cysteine; a, typically an aliphatic amino acid; X, C-terminal amino acid). GGTase-II attaches geranylgeranyl groups to two C-terminal cysteine residues in the Rab family of Ras-related GTPases. All three enzymes are heterodimers and catalyse bond formation with the involvement of an essential zinc ion^{1,5}.

Ideally, enzyme mechanistic studies include experimentally determined structures of each step along the reaction coordinate, but that is rarely possible in practice. Here we present two newly determined crystal structures of FTase complexes: one containing farnesylated Ras peptide product alone, **3**, and a complex that contains both the farnesylated peptide and an additional farnesyl diphosphate (FPP) substrate, **4**. Together with three previously determined structures (apo-FTase⁶, **0**, and complexes with FPP alone⁷, **1**, and nonreactive FPP analogue with unfarnesylated peptide^{8,9}, **2**), these new structures complete the structural analysis of the major steps along the reaction pathway that have also been defined kinetically¹⁰ (Fig. 1).

The transition state occurs between the steps represented by structures **2** and **3**. There is a 7.3 Å gap between the C₁ acceptor carbon in the FPP substrate and the attacking S _{γ} in the ternary reactant complex **2**. Unexpectedly, the structural rearrangements that we observe in complexes **0–4** are limited to localized structural rearrangements, involving only the ligands themselves and several directly interacting amino-acid side-chains. The two reactants could therefore be brought into juxtaposition not by major protein conformational changes, but by a rotation involving the first and second isoprene units (atoms C₁ to C₁₀) of the farnesyl moiety, resulting in a 5.7 Å movement of the C₁ atom towards the cysteine residue (Fig. 2a). This simple motion allows us to propose a structural model for the transition state (Fig. 2b) that is consistent with other mechanistic studies^{5,11–13}, and is anticipated to provide a guide for future theoretical and experimental studies.

The transition state is modelled as an intermediate conformation between the FPP structure observed in the ternary reactant complex **2** and product complex **3** (Fig. 2). The proposed structure retains elements of both the postulated electrophilic and nucleophilic components of reaction mechanism¹³. An important consideration in modelling the transition state is that product **3** is coordinated to the catalytic zinc ion, consistent with the proposed nature of the attack of the S _{γ} on C₁ (ref. 5). In the modelled transition state the β -phosphate retains the conformation observed in **2**, while rotation of the phosphodiester bond repositions the α -phosphate to permit interaction of the C₁ atom of the isoprenoid with the cysteine thiolate of the peptide. Lys 164 α and Tyr 300 β mutations decrease the rate of catalysis about 30–40-fold¹³. In the transition state Tyr 300 β , which forms a hydrogen bond with the β -phosphate in the ternary reactant complex **2**, now interacts with the bridging oxygen between the α -phosphate and the C₁ atom. This interaction may help stabilize the negative charge that develops on this oxygen atom. Lys 164 α may further stabilize the negative charge by forming a hydrogen bond with a α -phosphate oxygen. The zinc and the C β –S _{γ} bond of the cysteine thiolate orient a free pair of electrons for in-line nucleophilic attack on C₁ (Fig. 2b).

Magnesium, which is also required for optimal enzymatic activity¹⁴ and may coordinate the diphosphate moiety⁹, is modelled to interact with both the prenyl diphosphate and residue Asp 352 β .

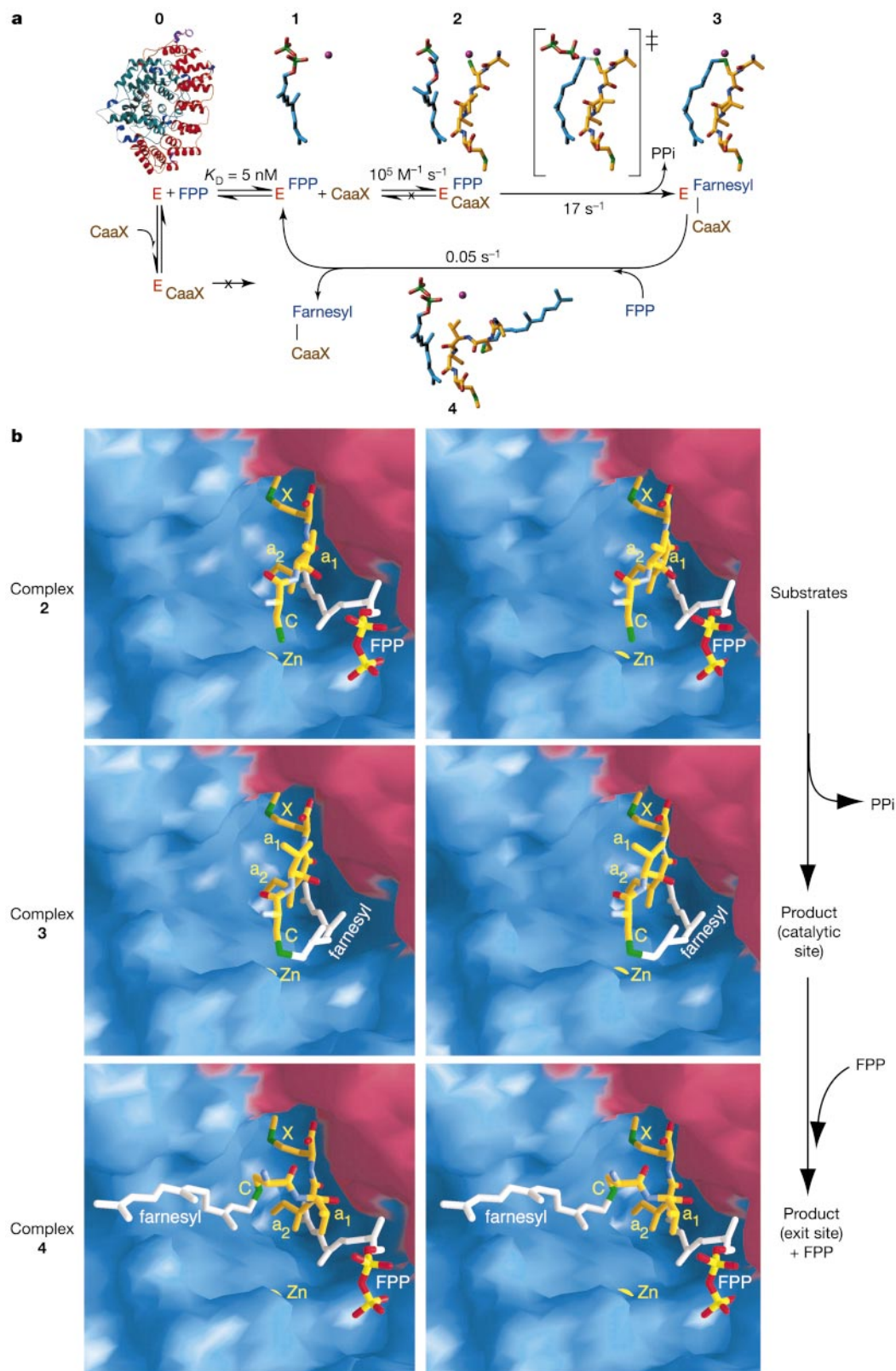


Figure 1 Structures along the FTase reaction path. **a**, Kinetic scheme¹⁰. **0**, Structure of unliganded FTase⁶ (PDB number 1FT1; red, α -subunit; blue, β -subunit). **1–4**, Observed ligand conformations in FTase crystal structures. **1**, Farnesyl diphosphate substrate (FPP, blue) bound to FTase⁷ (PDB number 1FT2). **2**, CaaX peptide substrate (yellow) and an analogue of FPP (blue), representing the ternary reactant complex⁹ (PDB number 1D8D). **3**, Farnesylated peptide product (farnesyl-CaaX), bound following catalysis. **4**, Both the product and FPP substrate bound following the addition of fresh FPP. Dissociation of the product regenerates complex **1**. The unliganded form may not exist *in vivo*. Double dagger

symbol, A modelled transition state along the reaction coordinate between **2** and **3**. The zinc ion is shown as a purple sphere. **b**, Complexes **2–4** shown (stereo) in the context of the molecular surface of the active site (red, α -subunit; blue, β -subunit; yellow, peptide; white, isoprenoid). In complex **4**, the isoprenoid of the product now occupies a shallow solvent-accessible groove. (For clarity, Lys 164 β , which coordinates the diphosphate of FPP, is omitted from the molecular surfaces. Only the CaaX motif of the peptides are drawn.) PP_i, inorganic pyrophosphate.

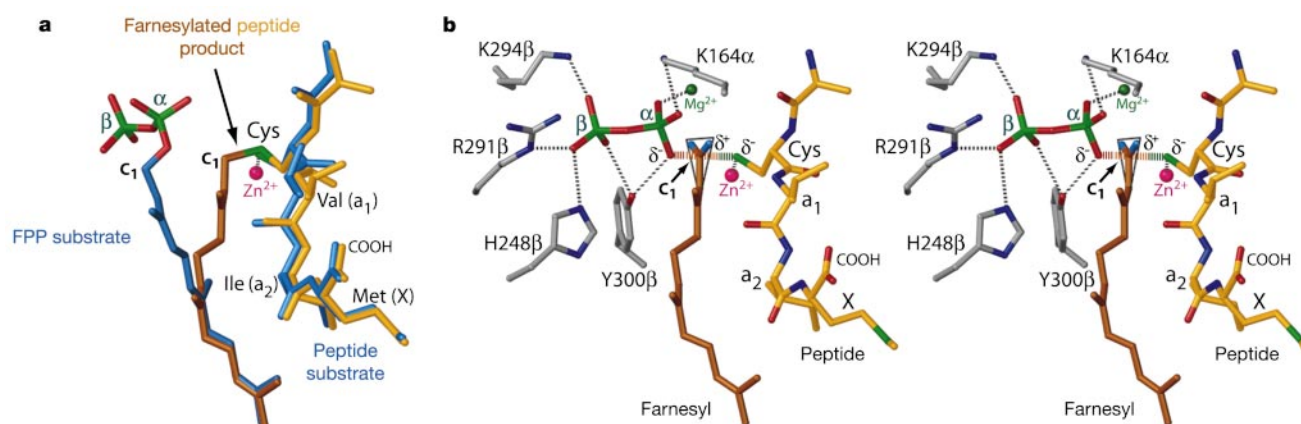


Figure 2 Comparison of reactant and product conformations and a proposed model for the transition state. **a**, Superposition of the observed structures of the substrates (blue) and the product (brown, isoprenoid; yellow, peptide) indicates reorganization of the isoprenoid upon product formation. The superposition was based on all FTase C_{α} atoms. **b**, Modelled transition state shown in stereo and in the same orientation as in **a**. The zinc ion activates the cysteine thiolate (green) for nucleophilic attack on the C_1 atom of the FPP

substrate. Dashed lines indicate the scissile and nascent bonds. The C_1 atom has trigonal bipyramidal geometry; a triangle indicates that the C_1 , C_2 , H_1 and H_2 atoms lie in a plane. The FTase diphosphate ligands (grey coloured residues) are in conformations observed in complex **2**. Residue Tyr 300 β stabilizes the developing negative charge on the bridging oxygen between the α -phosphate and the C_1 atom. Lys 164 α and magnesium also stabilize negative charge on the α -phosphate.

Unlike FTase, GGTase-I does not require magnesium¹. Sequence alignments¹¹ show that this residue is replaced by Lys 311 β in GGTase-I, suggesting that the positively charged lysine residue may substitute for the role of the magnesium ion in the reaction. To our knowledge, this is the first structure of a sulphur-alkylating zinc metalloenzyme complexed with its thioether product. This mechanism is likely to be conserved in other sulphur-alkylating enzymes, including GGTase-I, GGTase-II, methionine synthases, and the DNA repair enzyme Ada.

Formation of the ternary reactant complex is defined by structures **1** and **2**. Complex **2** reveals the unusual situation where one substrate (FPP) forms a substantial part of the binding surface for the other (CaaX peptide)^{8,9}, providing a structural basis for the observed ordered binding of the FPP and peptide substrates^{10,15}. In product complex **3**, the isoprenoid and CaaX portions form even more extensive van der Waals interactions with one another, in addition to contacts with FTase residues (Figs 2a and 3a). These interactions between the isoprenoid and peptide substrates therefore affect the mechanism for determining peptide specificity.

We note that product complex **3** is not the final step in the reaction. Kinetic studies have indicated that product release is slow for both FTase and GGTase-I^{10,16} and, at least in the case of FTase, requires addition of a substrate molecule¹⁷. This final step is represented by complex **4**, which reveals that binding of the fresh FPP substrate molecule moves the farnesyl group of the farnesylated peptide to a new binding site, the 'exit groove' (Figs 1b and 3b). This movement is accompanied by a conformational change in the CaaX moiety. There are therefore two CaaX peptide conformations in the product: extended (complex **3**) and type I β -turn (complex **4**). Nuclear magnetic resonance (NMR) studies of peptides bound to FTase observed a similar β -turn, which was postulated to correspond to a reactant rather than product complex^{18,19}. In complex **4**, the two C-terminal residues of the CaaX moiety occupy the same binding location as those in complexes **2** and **3**, which indicates that the peptide substrate is competitive with respect to the displaced product. The three C-terminal amino acids of the farnesylated peptide product make extensive van der Waals contacts with a new FPP substrate (Fig. 3b), suggesting that the amino-acid sequence of the CaaX motif may modulate product release. Multiple binding sites for product molecules have been observed in proces-

sive enzyme systems in which substrate binding is involved in product release or translocation (for example, the P and E sites of the ribosome²⁰). We note that complex **4** represents a direct observation of both substrate and product bound simultaneously to an enzyme that catalyses an apparently non-processive reaction.

Unlike FTase, the homologous GGTase-II enzyme is processive, since it adds two prenyl groups²¹ to its cognate protein substrates, the Rab proteins²², via a process that appears to proceed without dissociation of the mono-prenylated intermediate²³. Rab proteins are regulators of vesicular transport²². The controlled product release that we observe in FTase may provide an explanation for the mechanism of GGTase-II. The crystal structure of GGTase-II, which has been determined only in the absence of bound ligands,

Table 1 Data collection and refinement statistics

	FTase•product	FTase•FPP•product
Data collection (all data)		
Resolution range (Å) (outer shell)	50.0–2.1 (2.2–2.1)	50.0–2.2 (2.3–2.2)
No. of unique reflections/total	67,840/402,441	56,209/253,335
Mean I/σ_1	21.0 (2.8)	16.1 (1.8)
Completeness (%)	100	96.0
R_{sym} (%)	8.6 (56.9)	8.9 (56.0)
Space group $P6_1$, cell dimensions (Å), $a = b, c$	171.24, 69.36	171.20, 69.44
Refinement ($F \geq 2\sigma_F$)		
Completeness (%)	90.5 (76.8)	85.6 (58.3)
R -factor (%)	16.4 (23.5)	16.0 (25.5)
R_{free} (%)	20.2 (27.3)	20.4 (27.9)
No. of non-hydrogen atoms	6429	6471
No. of water molecules	421	428
r.m.s.d. bond lengths (Å)	0.008	0.007
r.m.s.d. bond angles (°)	1.4	1.4
Ramachandran plot (% residues) core, disallowed regions	92, 0	92, 0
B-factors (Å ²)		
Protein atoms	34.1	34.9
Product: CaaX, isoprenoid	29.6, 25.9	36.8, 49.5
FPP	-	26.7
Zinc	26.8	32.8
Water	39.8	42.3

$R_{\text{sym}} = (\sum (I - \langle I \rangle)) / (\sum \langle I \rangle)$, where $\langle I \rangle$ is the average intensity of multiple measurements. R -factor and $R_{\text{free}} = (\sum |F_o - F_c|) / (\sum |F_o|)$. R_{free} was calculated using the same ~5% of structure factors from each data set that were derived from those for the starting model, PDB 1D8D (ref. 9), and were held aside throughout the refinement. Parentheses indicate outer-resolution shell.

shows that most of the active site features are conserved with FTase²⁴. Biochemical studies have shown that, like FTase, GGTase-II has only one prenyl diphosphate binding site²⁵. In GGTase-II, however, we note a striking 7.5 Å diameter hydrophobic tunnel that originates from a shallow groove analogous to the exit groove in FTase and extends through the β -subunit (residues Tyr 30 β , Tyr 39 β , Arg 46 β , Gly 49 β , and Pro 288 β line the tunnel). The first geranylgeranyl isoprenoid attached to Rab may be sequestered in this tunnel following translocation of the mono-prenylated product from the site of catalysis (Fig. 4). In a manner analogous to that observed for FTase, this translocation may be associated with the binding of fresh geranylgeranyl diphosphate (GGPP) substrate^{23,26}. Furthermore, this translocation would allow the unmodified cysteine residue to interact with the zinc ion in preparation for the second catalytic reaction. This mechanism would explain why GGTase-II preferentially modifies Rab substrates that have one geranylgeranyl group already attached^{21,23}. In addition, the potential to insert different lengths of the geranylgeranyl moiety through the tunnel may explain how GGTase-II is able to modify cysteine residues at a variety of positions within a few residues of the C-termini of Rab substrates.

By capturing the FTase exit complex **4** we have identified a

previously unknown intermediate in the FTase reaction coordinate. This complex indicates that product does not dissociate simply from complex **3**, as previously supposed. The stability of complex **4** is not an artefact of crystal contacts, because these are absent in the active site. Complex **4** also persists in high concentrations of FPP substrate, and dissociates only if excess peptide substrate is added to the crystal. The biological role for this stable complex **4** is open to speculation. A portion of the CaaX moiety of the product in complex **4** occupies the same location as that of the protein substrate, so the ultimate dissociation of the product from the enzyme may be driven by binding of the next (unfarnesylated) peptide substrate. It is also possible that *in vivo* the complex of the enzyme and its farnesylated product persists until it encounters the enzyme responsible for cleavage of the aaX moiety²⁷, the next step in the post-translational processing of farnesylated proteins.

The structural details of the FTase reaction pathway presented here suggest that clinically active anti-cancer FTase inhibitors may have at least two primary modes of inhibition: blocking the peptide substrate site and, unusually, blocking the exit groove. Indeed, three structurally resolved complexes of FTIs bound to FTase have shown that this is the case. L-739,750 binds in a non-productive conformation at the peptide-binding site²⁸; the 3-aminopyrrolidinone

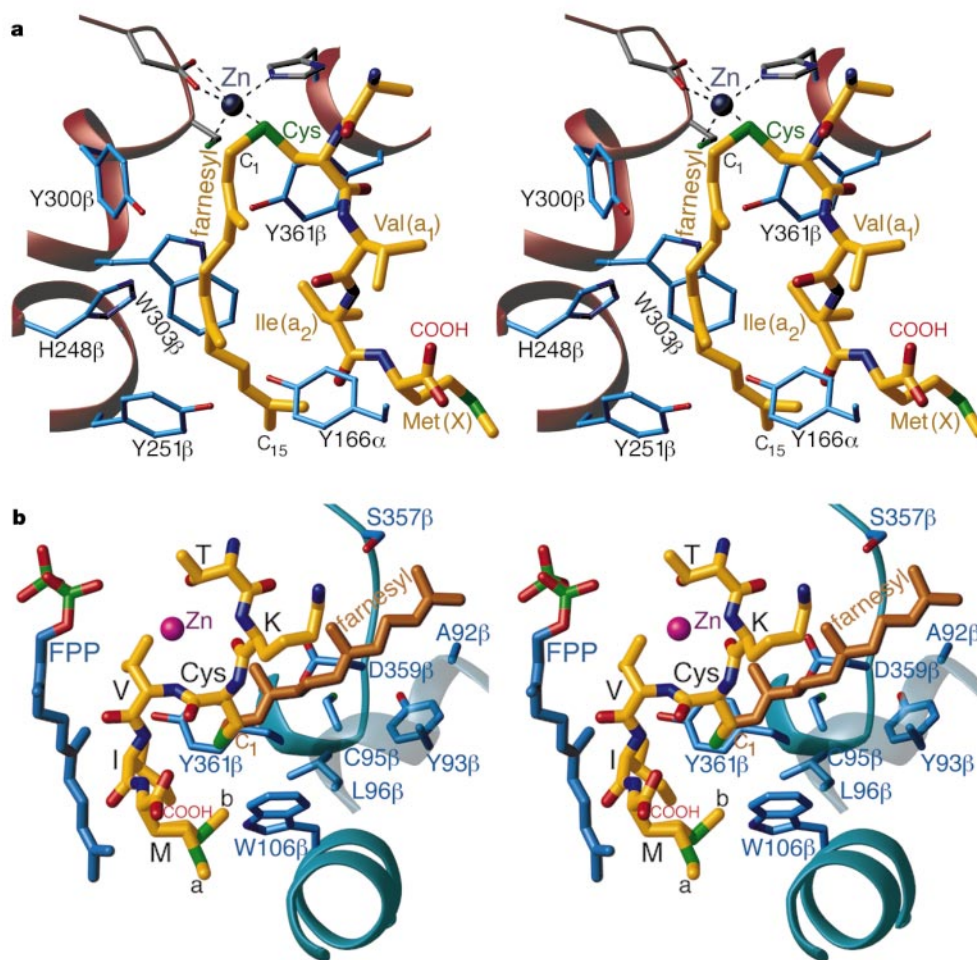


Figure 3 Farnesylated peptide product conformations at the active site (stereo). **a**, Product bound following catalysis (complex **3**). The zinc ligands (grey) and those residues that have van der Waals contacts with the first or second isoprene units of the farnesyl moiety (C₁ to C₁₀) are shown. Residues that interact with the third isoprene unit (C₁₁ to C₁₅) are the same as in complexes with FPP⁷ and FPP analogues^{8,9}.

b, Product and FPP bound simultaneously (complex **4**). Residues that interact with the farnesyl group of the product are shown. The two C-terminal amino acids of the product share the same conformation as observed in complexes **2** and **3**. The C-terminal methionine side chain of this product has alternate conformations, labelled 'a' and 'b'. Conformation 'a' is the orientation observed in complexes **2** and **3**.

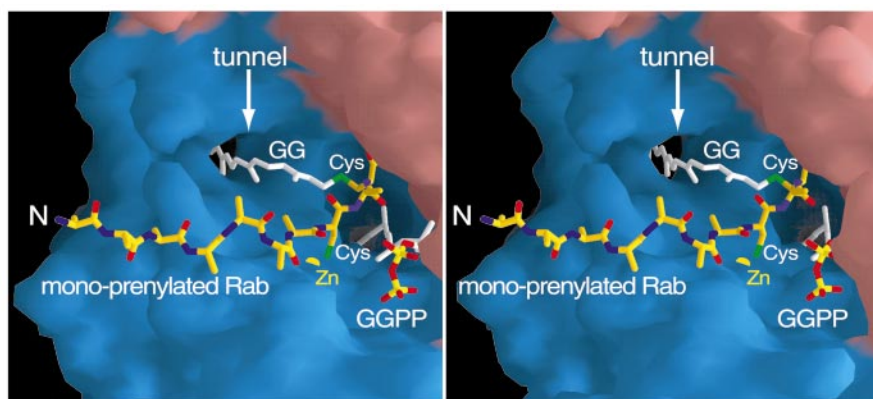


Figure 4 A model for processive prenylation catalysed by GGTase-II (stereo). The molecular surface of the active site of apo-GGTase-II (red, α -subunit; blue, β -subunit; PDB number 1DCE) is shown in the same orientation as that of FTase in Fig. 1. The step depicted here occurs before the addition of the second isoprenoid. Here, both mono-prenylated Rab and geranylgeranyl diphosphate (GGPP) are bound simultaneously to the enzyme. The geranylgeranyl isoprenoid (GG) of the mono-prenylated Rab peptide is modelled to occupy a tunnel (arrow) that extends from the site analogous to the FTase exit groove. GGPP is modelled in the conserved isoprenoid diphosphate binding site²⁴. This

organization of prenylated product and isoprenoid diphosphate substrate is analogous to FTase complex 4, with the modification that the zinc ion coordinates a second cysteine residue of the Rab substrate. This cysteine residue is now poised for the next catalytic reaction, analogous to the FTase reactant complex 2. The C-terminal sequence of the Rab peptide model is –CAC (in single amino-acid code), corresponding to that of Rab-3a. The location of the peptide substrate was derived from the binding of peptide observed in the FTase structures. The amino-terminal region of the α -subunit of GGTase-II (residues 1–20) is omitted from the molecular surface for clarity.

inhibitors also block the peptide site²⁹; and SCH 66336 (ref. 30) occupies both part of the substrate peptide site and the exit groove, reminiscent of product binding in complex 4. There is therefore ample opportunity for the design of additional drugs that are targeted to binding in the FPP, peptide or exit groove sites. These refined design strategies may become important in the design of drugs that specially target FTases of pathogens such as *Plasmodium falciparum* or *Trypanosoma brucei*, and do not bind to human FTase. The completeness of this high-resolution series of reaction path intermediates, and the observation that structural rearrangements along the reaction coordinate are limited to localized structural changes involving only the ligands and a few directly interacting amino-acid side-chains, provides an opportunity for theoretical mechanistic computations, assessment of ligand-docking algorithms, and structure-based drug design. □

Methods

Crystallization

Full-length recombinant rat FTase was expressed from Sf9 cells, purified, and concentrated to 16 mg ml^{−1} as described⁶. The product complex was prepared before crystallization by the following method: concentrated protein was incubated with FPP for 2 h on ice in order to form the FTase:FPP complex; this complex was incubated with KKKSKTKCVIM peptide for 30 min on ice to form the FTase:FPP:peptide complex; the reaction was initiated by the addition of required MgCl₂ to 5 mM final concentration and the mixture incubated at 20 °C for 30 min. The molar ratio for FTase:FPP:peptide used was 1:2:3. This mixture was used for crystallization under conditions described previously⁹. To obtain the complex with both FPP and product bound, a preformed product co-crystal was transferred to a solution of mother liquor supplemented with 10 μ M FPP for 20 h in a volume of 250 μ l, yielding an approximately 250-fold molar excess of FPP. Crystals of both complexes were transferred stepwise to cryoprotectant solutions⁹ that did not contain additional substrate or product molecules and were flash-cooled by plunging into liquid nitrogen.

Structure determination

Diffraction data were measured at 98 K with an R-axis IV image plate system mounted on a Rigaku RU-H3R generator. Data reduction and scaling were done using the DENZO and SCALEPACK programs. The crystals belong to the space group P6₁ and contain one complex per asymmetric unit. Coordinates for the ternary complex (PDB ID number 1D8D), which crystallized in the same space group with similar cell dimensions, were used to determine initial phases. Phase calculation and model refinement were done in X-PLOR using a bulk solvent correction. The final models contain α -subunit residues 55–377 and β -subunit residues 17–423. Table 1 contains a summary of diffraction data and refinement statistics; electron density for each product molecule is shown as Supplementary Information.

Received 13 February; accepted 24 June 2002; doi:10.1038/nature00986.

- Casey, P. J. & Seabra, M. C. Protein prenyltransferases. *J. Biol. Chem.* **271**, 5289–5292 (1996).
- Adjei, A. A. Blocking oncogenic Ras signalling for cancer therapy. *J. Natl Cancer Inst.* **93**, 1062–1074 (2001).
- Johnston, S. R. D. Farnesyl transferase inhibitors: a novel targeted therapy for cancer. *Lancet Oncol.* **2**, 18–26 (2001).
- Karp, J. E. *et al.* Current status of clinical trials of farnesyltransferase inhibitors. *Curr. Opin. Oncol.* **13**, 470–476 (2001).
- Huang, C., Hightower, K. E. & Fierke, C. A. Mechanistic studies of rat protein farnesyltransferase indicate an associative transition state. *Biochemistry* **39**, 2593–2602 (2000).
- Park, H.-W., Boduluri, S. R., Moomaw, J. F., Casey, P. J. & Beese, L. S. Crystal structure of protein farnesyltransferase at 2.25 Å resolution. *Science* **275**, 1800–1804 (1997).
- Long, S., Casey, P. J. & Beese, L. S. Co-crystal structure of protein farnesyltransferase with a farnesyl diphosphate substrate. *Biochemistry* **37**, 9612–9618 (1998).
- Strickland, C. L. *et al.* Crystal structure of farnesyl protein transferase complexed with a CaaX peptide and farnesyl diphosphate analogue. *Biochemistry* **37**, 16601–16611 (1998).
- Long, S. B., Casey, P. J. & Beese, L. S. The basis for K-Ras4B binding specificity to protein farnesyltransferase revealed by 2 Å resolution ternary complex structures. *Struct. Fold. Des.* **8**, 209–222 (2000).
- Furfine, E. S., Leban, J. J., Landavazo, A., Moomaw, J. F. & Casey, P. J. Protein farnesyltransferase: Kinetics of farnesyl pyrophosphate binding and product release. *Biochemistry* **34**, 6857–6862 (1995).
- Dolence, J. M., Rozema, D. B. & Poulter, C. D. Yeast protein farnesyltransferase. Site-directed mutagenesis of conserved residues in the beta-subunit. *Biochemistry* **36**, 9246–9252 (1997).
- Kral, A. M. *et al.* Mutational analysis of conserved residues of the beta-subunit of human farnesyl:protein transferase. *J. Biol. Chem.* **272**, 27319–27323 (1997).
- Wu, Z. *et al.* Farnesyl protein transferase: identification of K164 α and Y300 β as catalytic residues by mutagenesis and kinetic studies. *Biochemistry* **38**, 11239–11249 (1999).
- Reiss, Y., Brown, M. S. & Goldstein, J. L. Divalent cation and prenyl pyrophosphate specificities of the protein farnesyltransferase from rat brain, a zinc metalloenzyme. *J. Biol. Chem.* **267**, 6403–6408 (1992).
- Mathis, J. R. & Poulter, C. D. Yeast protein farnesyltransferase: a pre-steady-state kinetic analysis. *Biochemistry* **36**, 6367–6376 (1997).
- Yokoyama, K., Zimmerman, K., Scholten, J. & Gelb, M. H. Differential prenyl pyrophosphate binding to mammalian protein geranylgeranyltransferase-I and protein farnesyltransferase and its consequence on the specificity of protein prenylation. *J. Biol. Chem.* **272**, 3944–3952 (1997).
- Tschantz, W. R., Furfine, E. S. & Casey, P. J. Substrate binding is required for release of product from mammalian protein farnesyltransferase. *J. Biol. Chem.* **272**, 9989–9993 (1997).
- Stradley, S. J., Rizo, J. & Gierasch, L. M. Conformation of a heptapeptide substrate bound to protein farnesyltransferase. *Biochemistry* **32**, 12586–12590 (1993).
- Koblan, K. S. *et al.* NMR studies of novel inhibitors bound to farnesyl-protein transferase. *Prot. Sci.* **4**, 681–688 (1995).
- Cate, J. H., Yusupov, M. M., Yusupova, G. Z., Earnest, T. N. & Noller, H. F. X-ray crystal structures of 70S ribosome functional complexes. *Science* **285**, 2095–2104 (1999).
- Farnsworth, C. C., Seabra, M. C., Ericsson, L. H., Gelb, M. H. & Glomset, J. A. Rab geranylgeranyl transferase catalyzes the geranylgeranylation of adjacent cysteines in the small GTPases Rab1A, Rab3A, and Rab5A. *Proc. Natl Acad. Sci. USA* **91**, 11963–11967 (1994).
- Novick, P. & Zerial, M. The diversity of Rab proteins in vesicle transport. *Curr. Opin. Cell Biol.* **9**, 496–504 (1997).
- Thoma, N. H., Niculae, A., Goody, R. S. & Alexandrov, K. Double prenylation by RabGGTase can

- proceed without dissociation of the mono-prenylated intermediate. *J. Biol. Chem.* **276**, 48631–48636 (2001).
24. Zhang, H., Seabra, M. C. & Deisenhofer, J. Crystal structure of Rab geranylgeranyltransferase at 2.0 Å resolution. *Struct. Fold. Des.* **8**, 241–251 (2000).
 25. Desnoyers, L. & Seabra, M. C. Single prenyl-binding site on protein prenyl transferases. *Proc. Natl Acad. Sci. USA* **95**, 12266–12270 (1998).
 26. Thoma, N. H. *et al.* Allosteric regulation of substrate binding and product release in geranylgeranyltransferase type II. *Biochemistry* **40**, 268–274 (2001).
 27. Ashby, M. N. CaaX converting enzymes. *Curr. Opin. Lipidol.* **9**, 99–102 (1998).
 28. Long, S. B., Hancock, P. J., Kral, A. M., Hellinga, H. W. & Beese, L. S. The crystal structure of human protein farnesyltransferase reveals the basis for inhibition by CaaX tetrapeptides and their mimetics. *Proc. Natl Acad. Sci. USA* **98**, 12948–12953 (2001).
 29. Bell, I. M. G. *et al.* 3-Aminopyrrolidinone farnesyltransferase inhibitors: design of macrocyclic compounds with improved pharmacokinetics and excellent cell potency. *J. Med. Chem.* **45**, 2388–2409 (2002).
 30. Strickland, C. L. *et al.* Tricyclic farnesyl protein transferase inhibitors: crystallographic and calorimetric studies of structure-activity relationships. *J. Med. Chem.* **42**, 2125–2135 (1999).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com/nature>).

Acknowledgements We thank H. W. Hellinga for extensive discussions and K. L. Terry for assistance with crystallization and data collection. The work was supported by National Institutes of Health grants to L.S.B. and P.J.C., and an American Heart Association predoctoral fellowship to S.B.L.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to L.S.B (e-mail: lsb@biochem.duke.edu). The coordinates for the complexes with farnesylated product and with both FPP and product, as well as the transition state model have been deposited in the Protein Data Bank (Research Collaboratory for Structural Bioinformatics), PDB ID numbers 1KZP, 1KZO and 1KZR, respectively.

Array for proteins, array for DNA

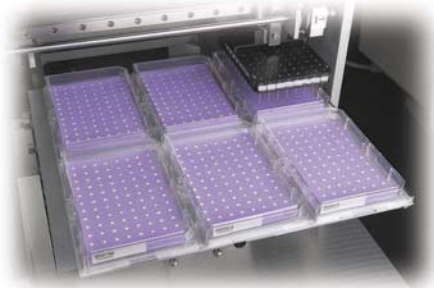
Microarrays, DNA libraries and sequencing — and siRNAs join the fray.

IsoCode I.D. library card

Schleicher & Schuell www.s-and-s.com

Ambient storage of DNA

IsoCode I.D. is a DNA library card for ambient archiving of DNA libraries. It is made of a chemically treated paper that is bactericidal, fungicidal and virucidal, keeping the DNA in biological samples that are dried on it safe from degradation. The chemistry also inactivates DNases, allowing indefinite storage of DNA at ambient temperatures. A colour indicator facilitates identification of the location of a sample after application on the paper. The card fits into a Nunc OmniTray and is compatible with most robotic and manual multi-tip applicator platforms.



Card carrying: Schleicher & Schuell's IsoCode I.D.

Atlas

BD Biosciences www.bdbiosciences.com

Ready-to-print cDNA and long oligos

BD Atlas provides a complete and tested collection of named array targets for use in printing arrays. The company offers cDNAs and long oligos (approximately 80 bases) for printing glass, plastic or nylon arrays.

TranSignal II

Panomics www.panomics.com

Protein/DNA array for multiple transcription factors

The second edition of the TranSignal protein/DNA array from Panomics performs simultaneous, high-throughput, functional analysis of up to 96 eukaryotic transcription factors (TFs). It can be combined with the company's TranSignal I, which includes 54 TF consensus sequences, to study a total of 150 unique TFs. The procedure for both arrays is simple and straightforward, requiring no special equipment for detection. All reaction and detection reagents are included in the kit.

BigDye Terminator Cycle Sequencing Kits

Applied Biosystems www.appliedbiosystems.com

Updated formulations to meet most sequencing needs

Applied Biosystems has introduced two new sequencing formulations designed to provide researchers with longer read lengths and higher success rates in sequencing difficult templates. BigDye Terminator version 3.1 is suitable for most sequencing applications, including *de novo* sequencing and resequencing. Version 1.1 is designed for speciality applications such as sequencing PCR products associated with comparative sequencing when using a rapid instrument run.

TraXis

Micronic BV www.micronic.com

Coded tubes make an indelible impression

The TraXis coded tube system is designed to eliminate the possibility of false identities and ensure a secure sample audit trail. An injection moulding technique produces a unique two-dimensional code. The laser encryption code at the bottom of each tube cannot fall off or be wiped off and is resistant to mechanical and chemical abrasion. Tubes may be encrypted with up to 8.5 billion different combinations of a numerical code. TraXis readers can scan all 96 tubes in a tube rack without the need for tube removal. Single-tube scanners are also available for integration into an automated system.

Cancer siRNA Oligo Set

Qiagen www.qiagen.com

They are small, and interfere

Qiagen's Cancer siRNA Oligo Set is, the makers say, the first set of disease-specific siRNAs for the life sciences market. The kit includes two siRNAs to each of 139 clinically

and scientifically relevant cancer genes. Each siRNA has been developed with maximized gene-silencing potential in mind. The Cancer siRNA Oligo Set provides a method by which a large number of cancer-related genes can be studied at once. The siRNAs are 21-nucleotide sense and antisense RNA oligonucleotides, annealed and purified to >97% purity by HPLC.

Q-DIS/WIRELESS

Creon.Lab.Control

www.creonlabcontrol.com

No plugs for this device

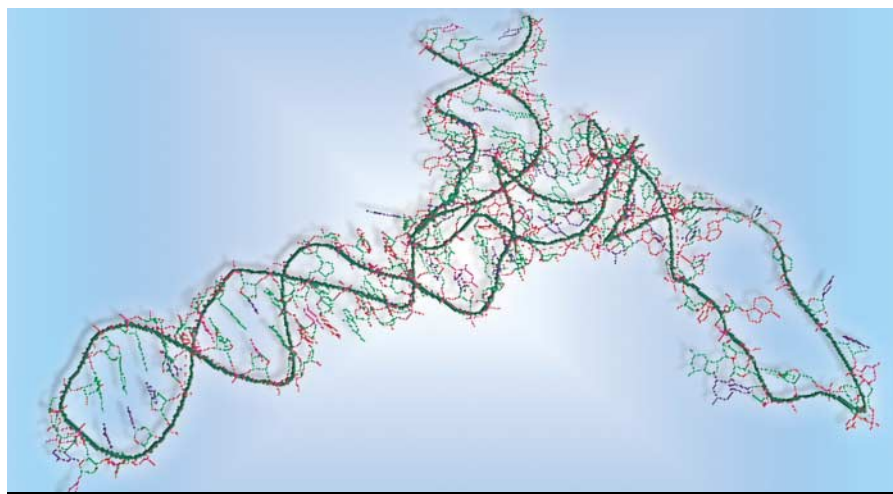
This web technology-based, wireless data transfer tool supports all hand-held computers equipped with wireless LAN network adapters compatible with IEEE 802.11b (Wi-Fi). Data are entered and automatically sent to the Q-DIS/WIRELESS server, and from there directly into the Q-DIS, QM LIMS or Q-DIS/R analytical workflow system. The technology can also be integrated into existing LIMS systems. Data are protected against loss by being transferred immediately, without being stored on the hand-held computer. The data transfer is documented by a fully traced data trail.

Vacuum prep tool

Pyrosequencing AB www.pyrosequencing.com

Speed up DNA sample preparation

Pyrosequencing's hand-held vacuum prep tool uses the speed of vacuum filtration to prepare a batch of 96 DNA samples for genetic analysis in less than 15 minutes. Pipetting is minimized and the filter probes within the tool are designed for repeated use, with each set able to prepare over fifty 96-well plates.



Cancer siRNA: Qiagen's new RNA set covers a broad range of cancers.

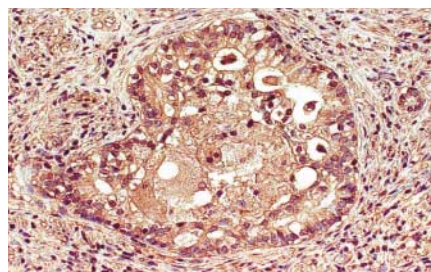
GeneXPRO

InnoGenex

www.innogenex.com

Gene expression service

GeneXPRO is a contract service for gene expression profiling, target localization and validation that will perform assays to analyse cellular and tissue-level expression. The gene profiling service includes design, synthesis and labelling of probes, correlation of gene expression profile with normal and pathogenic conditions on the cells, tissues and tissue arrays and determination of changes in gene expression in response to candidate drugs. Assays include *in-situ* hybridization (ISH). Translation of transcripts into specific proteins can be localized and analysed by immunohistochemistry (IHC). Screening of mouse monoclonal antibodies on mouse tissues, IHC and ISH on human and rodent tissues, single and multiplex staining procedures, microwave target retrieval and special tissue pre-treatments are also offered.



GeneXPRO in action — showing β -actin *in situ*.

gels during the denaturing gel electrophoresis process. The marker is made up of eight highly purified proteins of known molecular weights (lysozyme, trypsin inhibitor, carbonic anhydrase, ovalbumin, BSA, phosphorylase B, β -galactosidase and myosin) and is suitable for both high and low MW applications. It is supplied as a ready-to-load, pre-mixed solution and covers a MW range of 14–212 kilodaltons, producing eight distinct bands when run.

Apo E4, Apo H, Apo SAA, HDL, LDL, Lp(A) and VLDL. Purified mouse Apo AI and rabbit Apo E are also available.

Zenith 340

Anthos Labtec Instruments

www.anthos-labtec.com

Absorbing reading

This microplate reader from Anthos features multiple measurement modes for mono- and bichromatic readings as well as kinetic and position scanning readings, and is compatible with plate formats from 6 to 384 wells. The stand-alone model can be used with a Windows CE computer and fully integrated into local computer networks. A remote-control version of the Zenith 340 comes with a data acquisition package that is compatible with many data analysis tools in addition to the ADAP software from Anthos. It is suitable for absorbency reading applications in life sciences as well as for routine ELISAs.

PfuUltra

Stratagene

www.stratagene.com

High-fidelity DNA polymerase

Stratagene claims that its *PfuUltra* DNA polymerase, which is formulated using a new, genetically engineered mutant of *Pfu* DNA polymerase, increases accuracy by threefold compared with other thermostable polymerases. *PfuUltra* incorporates Stratagene's ArchaeMaxx polymerase-enhancing factor to prevent poisoning of archaeal polymerase reactions and facilitate robust PCR product yield and amplification of long genomic targets. Applications include PCR cloning, site-directed mutagenesis and RT-PCR, where incorporation accuracy is vital.

UnBlot

Pierce

www.piercenet.com

Transfer-free in-gel chemiluminescent detection

The UnBlot kit for biotinylated antibody probes uses a transfer-free in-gel detection method, reducing handling time from days to hours and saving materials and equipment costs. Because it allows researchers to view the target without having to transfer it to a membrane, the UnBlot kit eliminates skewed results caused by low molecular weight (MW) proteins transferring more efficiently than higher MW proteins. The kit, which is sensitive to 1 ng, uses a streptavidin-based system, resulting in less nonspecific binding than that which occurs with secondary antibodies. The kit is compatible with stripping and reprobing protocols as well as protein staining.

CAMPATH-1

Serotec

www.serotec.co.uk

CD52 antigen

The CAMPATH-1 antigen is designated as CD52, a small, GPI-linked molecule that is strongly expressed by lymphocytes and monocytes, but much more weakly by granulocytes. Antibodies against CD52 have been used for immunotherapy in bone marrow transplantation and autoimmune disease and, more recently, in the treatment of leukaemia and lymphoma. CAMPATH-1M and CAMPATH-1G are available from Serotec for use in research studies of CD52 expression and function. In addition to the purified IgM antibody, the IgG antibody is offered in FITC and RPE conjugated formats that are suitable for the assessment of CD52 expression using flow cytometry.

Protein MW Marker

Amresco

www.amresco-inc.com

Wide-range, pre-mixed marker

The Protein Molecular Weight (MW) Marker from Amresco is used to establish the apparent molecular weights of proteins run on sodium dodecyl sulphate (SDS) polyacrylamide

Apolipoprotein antibodies and proteins

Biodesign

www.biodesign.com

A is for Apo

Biodesign's line of apolipoprotein products includes antibodies specific for Apo (a), Apo AI, Apo AII, Apo B, Apo C, Apo CI, Apo CII, Apo CIII, Apo E, Apo H, Apo J and Lp(a). Depending on the monoclonal or polyclonal antibody chosen, antibody formats available are neat serum, purified unconjugated antibody and/or conjugated to biotin or peroxidase. Antibody host species include mouse, goat, hamster, rabbit and sheep. Although most antibodies are anti-human specificity, some anti-mouse, anti-hamster and anti-rabbit products are available. Applications include western blot, ELISA, competitive EIA, immunodiffusion assay, enzyme conjugation or biotinylation. Purified human proteins include Apo AI, Apo AII, Apo B, Apo CI, Apo CII, Apo CIII, Apo E, Apo E2, Apo E3,

These notes are compiled in the Nature office from information provided by the manufacturers. Press releases to be considered for publication can be sent to newproducts@nature.com.



Marker time: gel calibration, from Amresco.

Contacts

Publisher: Ben Crowe

Editor: Paul Smaglik

Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building

4 Crinan Street

London N1 9XW, UK

Tel +44 (0) 20 7843 4961

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Senior European Sales Manager:

Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)

Andy Douglas (4975)

Frank Phelan (4944)

Netherlands/ Italy/ Iberia:

Evelina Rubio Hakansson (4973)

Scandinavia: Silje Opstrup (4994)

France/ Belgium:

Amelie Pequignot (4974)

Production Manager: Billie Franklin

To send materials use London
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Hind Berrada (4935)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Office

Germany/ Austria/ Switzerland:

Patrick Phelan, Odo Wulffen

Tel + 49 89 54 90 57 11/-2

Fax + 49 89 54 90 57 20

e-mail: p.phelan@nature.com

o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,

10th Floor, New York, NY 10010-1707

Tel +1 800 989 7718

Fax +1 800 989 7103

e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

US Advertising Coordinator:

Linda Adam

Japan Head Office, Tokyo

MG Ichigaya Building (5F),

19-1 Harakutamachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

e-mail: k.johnson@naturejpn.com

Asia-Pacific Advertising Manager:

Keyyn Johnson

naturejobs

Alternative roads to success

An analyst, an editor and a business-development professional. Three different professions, one striking similarity — they all got their current jobs without doing a postdoc. They revealed this to an alternative-career roundtable at a recent career fair co-sponsored by *Naturejobs* and the New York Biotechnology Association.

Eric Schmidt, managing director of SG Cowan, a financial analysis company in New York, realized midway through his PhD at the Massachusetts Institute of Technology that he didn't want a career in "bench science". His avid interest in the stock market led him to some business courses and his job at Cowan.

Meeghan Sinclair, an associate editor on *Nature Biotechnology*, wasn't sure she was ready to focus narrowly on one problem. "I wanted to be involved in learning about science but have a broader experience of science," she says. Her present position allows her to do just that, and also to be involved in technologies that have direct applications.

Catharine Johnson, manager of business development for Regeneron, a biotech company in New York, knew even before she entered university that an academic career probably wasn't for her — she had enjoyed a high-school internship with a biotech firm. Her extracurricular activities as a graduate student, including organizing panels on alternative careers, further piqued her interest.

All three rate their jobs as satisfying, but that doesn't mean that advancing their careers has been easy. Analysts' jobs are tied to a volatile stock market, editors' to a relentless cycle of deadlines, and business-development experts' to bridging the equally uncertain worlds of science and commerce. But for all three, such obstacles were better than the alternative — a series of postdocs and the anticipation of tenure track.

Paul Smaglik
Naturejobs editor



Contents

SPECIAL REPORT

Plant biology goes back
to its roots

p4

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

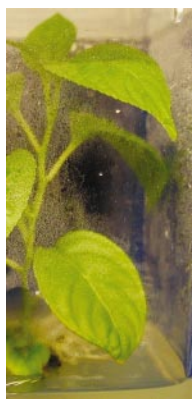
FOCUS

SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS



While industrial agrobiotech R&D falters, opportunities in plant biology in the public sector are growing, says Virginia Gewin.

Anita Brinker's career path is a sign of changing times in plant science. Formerly an agricultural chemist in multinational giant DuPont's crop-protection division, she, like 450 of her colleagues, was a victim of corporate downsizing three years ago. She now works at Rutgers University in New Brunswick, New Jersey, putting her chemistry background to use on a potential plant-based drug. This industry-to-academia

route may become more familiar as corporate research dollars dry up and university science renews its commitment to basic plant biology.

The downsizing of DuPont's crop-protection division was part of a brief 'life-sciences' trend in the 1990s to amalgamate pharmaceutical and agrobiological research and development.

A number of large drug and agricultural biotechnology companies merged, anticipating synergies from sequenced genomes and advances in biotechnology. But the benefits failed to materialize and the trend proved short-lived. After the conglomerates failed to meet market expectations, many divested themselves of their recently acquired agricultural interests.

Although the 1990s expansion concentrated on using transgenics to manipulate 'input traits' — such as resistance to pests, diseases and herbicides — Brinker, like many other plant researchers, is now focusing her energy on discovering and exploiting existing valuable plant traits, known as 'output traits'. Ilya Raskin, a plant biochemist at Rutgers University and Brinker's boss, even thinks that input-trait research — the bread and butter of traditional agricultural business — is passé. Growing more calories cheaply isn't the future as he sees it. "How do we grow health and wellness? That's where jobs should grow," he says.

It is not just that there are more jobs around — whole new fields are being conceived as technology continues to spur marriages between disciplines. Metabolomics, or the complete metabolic profiling of cell function, is one of these new unions. And the Nutraceuticals Institute at Rutgers University recently hired Mohamed Rafi, originally from India, as its first professor of nutrigenomics — the new discipline that develops functional foods with specific health benefits.

With the plethora of new research avenues, the most sought-after employees are those with multiple skills. "If you want to make a pharmaceutical out of a plant sample, you want people who know plant biology as well as chemical engineering. These cross-disciplinary people are really hard to find," says Charles Arntzen, a pioneer in the development of 'edible vaccines' at Arizona State University in Tempe.

Not surprisingly, cross-disciplinary training is a top goal for phase 2 of the \$400 million US National Plant Genome Initiative (NPGI). The directors of the programme foresee a need to train people in disciplines not usually associated with plant biology, such as computer science, mathematics and chemistry.

It is something of an irony that public programmes such as the NPGI were started to redress the perceived imbalance between the public and industrial research effort in plant biology, and are now thriving after the industrial efforts that triggered them have tailed off (see 'Taking the public initiative', opposite).

SUSTAINED EFFORT

One new player on the public scene is the Donald Danforth Plant Science Center, a \$75-million not-for-profit research centre that opened in 2001 in St Louis, Missouri. It intends to target the needs of developing nations and, ultimately, to begin a revolution in plant science that will lead to a more sustainable world. It will tackle research problems that are likely to be overlooked by industry. Projects that yield small profit margins won't be done by industry, says Roger Beachy, president of the centre. Beachy, who developed the world's first genetically altered crop plant — a virus-resistant tomato — is currently filling the remaining research positions at the centre. At capacity, the centre will employ 300 researchers and administrators.

Academia is really shoring up the overall investment in plant research, agrees John Hamer, chief scientific



Growth industry: Anita Brinker (above), Ilya Raskin (right) and Charles Arntzen (below left) with colleague Richard Mahoney.



RUTGERS UNIV.

RUTGERS UNIV.

ARIZONA STATE UNIV.

officer of Paradigm Genetics of Research Triangle Park, North Carolina, a company that applies genomics to both drug discovery and agriculture.

Although industry hiring remains stagnant, the larger companies are following academia's lead in turning their attention to output traits as a route to new products. "The field of output traits is more open, a new marketplace," says Jim Bailey, global human-resources leader for research and development at Dow AgroSciences in Indianapolis. Potential products include sunflowers that could produce more oil, and tomatoes with higher levels of the cancer-preventing agent lycopene. Tweaking output traits has an added benefit — it doesn't necessarily involve transgenics, thus avoiding regulatory constraints and consumer acceptance issues.

THE WHOLE PICTURE

"We need people who've got some whole-organism knowledge to set detail into context," says David Lawrence, head of research at Syngenta in Basel, Switzerland. Basic plant physiologists and cell biologists are as much in demand as people who specialize in computational biology, protein structure and protein function, he says. Lawrence considers that input-trait research will continue, but it will have new aims, such as finding fresh targets for chemical pesticides, engineering crops for environmental challenges, and putting nitrogen fixation into a wider range of crop plants.

Start-up companies are most likely to be the output-trait innovators, if they can build partnerships in the current economic climate. Paradigm Genetics, a proven pace-setter, is successfully upholding the 'life-sciences' approach. As a pioneer in bioinformatics, functional genomics and metabolomics, it provides companies with genomic technology to create new products for both agriculture and human health.

Although the big conglomerates gave up on the life-sciences approach, the synergies may still be there. New techniques are rapidly opening up a number of research directions. "It's hard to explain the speed and scale of projects we undertake compared with even five years ago," says Caroline Dean, who studies the control of plant flowering time at the John Innes Centre in Norwich, UK.

Whatever the sector, the science will drive the jobs. "I think this is, from a science standpoint, one of the most exciting times in plant biology," says Roger Wyse,

Taking the public initiative

Funding programmes for plant biology are flourishing globally, and driving jobs. The \$400-million US National Plant Genome Initiative (NPGI) is one of the biggest. It is supporting the sequencing of each of the agronomically important families of Poaceae (grasses), Fabaceae (legumes) and Solanaceae (potato/tomato). Maize will probably be added. The genome of *Populus* (poplar) — a model species in forestry — is currently being sequenced by a joint US, Canadian and Swedish effort.

Other international collaborations are also stimulating the field. The *Arabidopsis* Functional Genomics Network is a multinational effort designed to uncover the

function of the estimated 28,000 *Arabidopsis* genes. The *Arabidopsis* 2010 programme is the US National Science Foundation's \$36-million contribution to that goal.

European partnerships in particular are flourishing. The US\$200-million French plant-genomics programme Genoplante, a collaboration between industry, government and academia, seeks to improve regionally important crops such as peas, rapeseed and sunflower seeds. In Germany, the government sponsors 90% of the 50-million-euro (US\$49-million) budget of the Genome Analysis of the Plant Biological System. V.G.

Sowing the seeds: the Donald Danforth Plant Science Center (left) is planning a revolution in plant science. And at Rutgers University, researchers are investigating *Tripterygium wilfordii* (right) as a potential source of an anti-arthritis drug.

RUTGERS UNIV.

a managing director at Burrill & Co., a San Francisco-based venture-capital firm. A diverse background and flexible attitude will serve up a number of career options. That approach helped Brinker to move from industry to academia, and she is not alone in surviving the life-sciences fallout. "The people that I've kept in touch with nearly all found jobs within a year," she says. And with an estimated 10,000–100,000 plant compounds still to be discovered, it seems that plant biotechnology may be a fruitful avenue to pursue. ■

Virginia Gewin is a freelance science writer based in Corvallis, Oregon.

Mohamed Rafi (left) and colleagues hope to develop functional foods that have specific health benefits.

